Higher-Order Methods for Solving Maxwell's Equations in the Time-Domain

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

von der Fakultät für Physik der Universität (TH)
Karlsruhe

genehmigte

DISSERTATION

von

Dipl. Phys. Jens Niegemann
aus Saarbrücken

Tag der mündlichen Prüfung: 16.01.2009
Referent: Prof. Dr. K. Busch
Korreferent: Prof. Dr. M. Wegener

Contents

1 Chapter 1

Introduction to Nanophotonics

The fascination of light is probably as old as mankind itself. Already the Greek philosophers Empedokles, Plato and Aristotle were debating over what causes the colors and how light and colors are related [1]. Over the course of the following 2500 years, hundreds or even thousands of people contributed to our modern understanding of light. Some of the more recent milestones on this route were the formulation of Maxwell's equations (1865) [2] and Einstein's concept of the photon (1905) [3]. In the 1940s, the development culminated in the formulation of quantum electrodynamics (QED), which allows a consistent description of light and its interaction with matter. While the field of QED and quantum optics is undeniably fascinating, in this thesis we will concentrate on systems where light can be described as a classical electromagnetic wave.

Despite the fact that Maxwell's theory is almost 150 years old, it still poses significant theoretical challenges. While analytic solutions can be found for certain simplified systems, there exists no universal method to exactly predict the behavior of light in arbitrary systems. Over the past decades, this shortcoming became particularly apparent when fabrication technology advanced to a point where it is possible to create almost arbitrary geometrical structures on the nanometer scale. As it turns out, these nanostructures expose a rich variety of qualitatively new physical effects, thereby allowing us to gain unprecedented control over light propagation and light-matter interaction. Some of these effects have sparked the imagination of scientists, leading to futuristic visions ranging from single-photon all-optical computing [4] to cloaking in the visible spectrum [5]. In addition, there is a variety of practical problems where nanophotonic devices might present a solution.

1.1 Integrated Photonics and Dielectric Resonators

Some of the most promising near-term applications can be found in data communication, where light already plays a dominant role. Almost the entire backbone of the modern telecommunication networks is based on optical fibers. While the transport of information is well under control, it is the processing of the data that still presents major challenges. Today, these processing tasks are usually performed by electronic

(a)

(b)
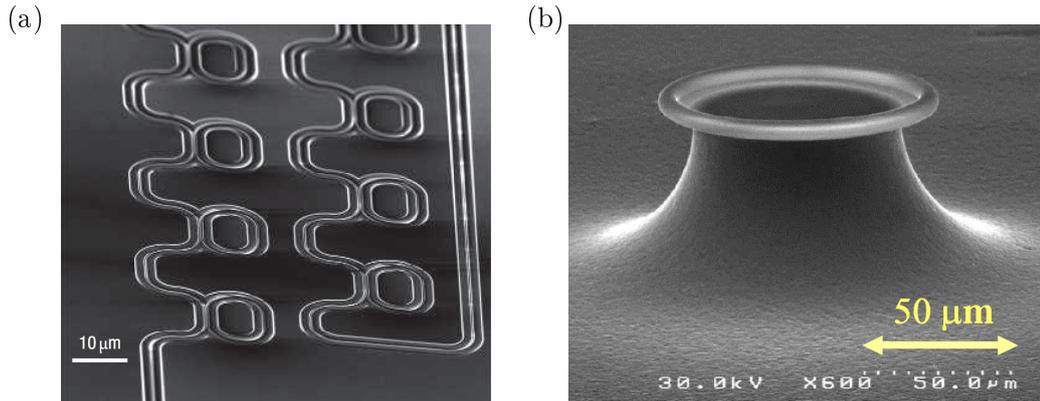
Figure 1.1: *Electron micrographs of two different realizations of dielectric resonators: (a) An integrated photonic circuit with ring resonators (adapted from Ref. [6]). (b) A toroidal microresonator used for single-molecule detection (adapted from Ref. [8]).*

devices, so that information first needs to be converted from light to electric signals and, afterwards, the processed data must be converted back again. Here, integrated nanophotonic devices which allow to filter, delay or switch signals all-optically could deliver significantly higher data rates at reduced cost and with less energy consumption.

Recently, some progress was made in that direction, when researchers at IBM presented realizations of compact optical buffers [6] and of high-throughput optical switches [7] on a chip. Both devices are based on nanophotonic ring-resonators similar to those depicted in Fig. 1.1(a). Particularly impressive in this work is the enormous precision achieved in manufacturing those integrated photonic devices. For the delay line shown in Fig. 1.1(a), the authors reported a mean surface roughness of only 1.1 nm, while the rings have a radius of 6.8 μm.

Besides their applications in optical data processing, there is a second fascinating field, where optical resonators truly excel. In recent experiments [8], it was demonstrated that one can utilize photonic microresonators as ultra-sensitive biological sensors. Employing a particular resonator structure (see Fig. 1.1(b)), the authors managed to reliably detect single biological molecules. It should be noted that sensing and information processing are only two examples out of the repertoire of possible applications of dielectric resonators. They can also be used to enhance nonlinear effects [9] or to study quantum-electrodynamical effects [10]. Furthermore, microresonators were recently considered for investigating optomechanical interactions [11].

What most applications of dielectric resonators have in common is that the structures need to be manufactured to very high precision. Consequently, this accuracy must also be reflected in corresponding theoretical calculations if reliable predictions or tangible design proposals are desired. It is this problem of modeling nanophotonic structures with high precision that will be the main topic of the present thesis.
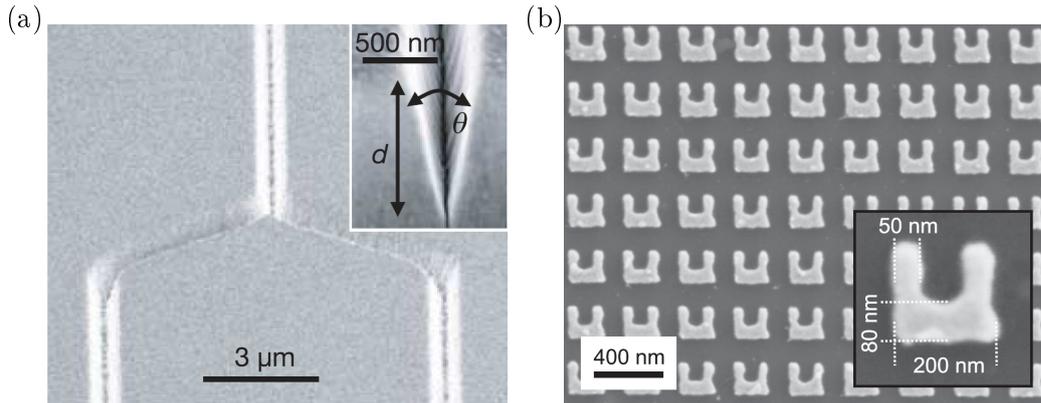
Figure 1.2: *Typical examples of metallic nanostructures: (a) A plasmonic channel waveguide splitting (adapted from Ref. [14]). (b) Electron micrograph of a metamaterial consisting of tiny split-ring resonators (adapted from Ref. [15]).*

1.2 Plasmonics and Metamaterials

Besides dielectric microresonators, the past years also witnessed an upsurge of interest in metallic nanostructures. Traditionally, besides their application as mirrors, metals did not play a major role as optical materials. It was only in the past decade that researchers realized how nanostructured metals can provide unique ways of manipulating light at length scales much smaller than its wavelength.

One of the pivotal experiments in this direction was performed by T. Ebbesen and coworkers in 1998 [12]. They manufactured a very basic array of subwavelength holes in a thin metallic film and measured the transmission. For certain wavelengths, the transmission exceeded theoretical estimates by several orders of magnitude. This result demonstrated clearly, that the traditional theory of optics often breaks down in the context of nanostructures. To properly understand the observed effects, one has to rigorously solve the fundamental equations governing light propagation, i.e., Maxwell's equations.

Today, there is a whole field of research focused on exploring the optical properties of metallic nanostructures. This area is often termed "plasmonics" and, within a few years, this young field has already brought forward a number of exciting ideas and discoveries. One of the key points of plasmonic structures is that they allow to concentrate electromagnetic fields in very small volumes. Thereby, for certain applications, they effectively allow to overcome the diffraction limit. Potential applications of plasmonic devices are manifold and range from the improvement of spectroscopy [13] to signal transport on very small length scales [14] (see Fig. 1.2(a)).

Closely related to plasmonics is the field of optical metamaterials. These man-made media consist of periodic arrangements of metallic nanostructures (see Fig. 1.2(b)) which should ideally be much smaller than the wavelength of the incoming light. If this

is the case, the light is not able to resolve the individual scatterers and instead "feels" an effective medium. By varying the shape, orientation and material of the building blocks, it becomes possible to tailor the optical properties of the metamaterial. In particular, it was demonstrated that one can achieve a material with negative refractive index [16] for visible light.

As alluded to above, an accurate theoretical description of metallic nanostructures mandates a rigorous solution of Maxwell's equations. Here, a particular problem is the effect that a sharp metallic corner can lead to a singularity in the electric field. This phenomenon—known as the lightning-rod effect—can, on the one hand, be exploited to create strongly enhanced electric fields. On the other hand, it massively complicates the accurate theoretical treatment, since one has to deal with divergent fields.

1.3 The Role of Numerics in Nanophotonics

As mentioned previously, exact analytical solutions to Maxwell's equations are only known for a few canonical structures such as spheres, ellipsoids or infinitely long cylinders. In all other cases, one usually has to resort to numerical methods. As long as a system only contains linear media, there are two basic approaches one can follow. Either one pursues a time-domain solution or one applies a Fourier transform and solves Maxwell's equations in the frequency-domain. For the latter class of problems, sophisticated finite-element methods (FEMs) are available which allow to accurately model complex geometries. However, in many cases a time-domain approach is preferable, because it allows to investigate the propagation of short pulses and transient effects more naturally. In addition, a time-domain solver is the only viable choice if the system contains any form of nonlinearity.

When researching time-domain methods with respect to Maxwell's equations, one inevitably comes across the celebrated finite-difference time-domain (FDTD) method. This scheme was originally proposed in 1966 by Kane Yee and has since become the primary means to computationally model the propagation of electromagnetic waves. The beauty of this method lies in its simplicity and versatility. By employing a rectangular grid and low-order discretizations, one obtains an explicit update scheme which can be implemented with only a few lines of computer code. This allows the scientist to focus on the research instead of spending time and energy on the implementation of a numerical scheme. The undeniable importance of the FDTD method is witnessed by the fact that more than 12,000 scientific articles[1] investigating or employing this method were published over the past 40 years.

However, for complex systems containing curved surfaces and material interfaces, the unadapted grid and the low-order approximations can also present a severe weakness (see Fig. 1.3). In these cases, the required computational effort of the FDTD method increases dramatically and one often needs large compute clusters with enor-

---

[1]The number was obtained by a search in the ISI database [17] for articles with either "FDTD" or "finite-difference time-domain" occurring in the title or abstract. Thereby, it can only serve as a lower limit.
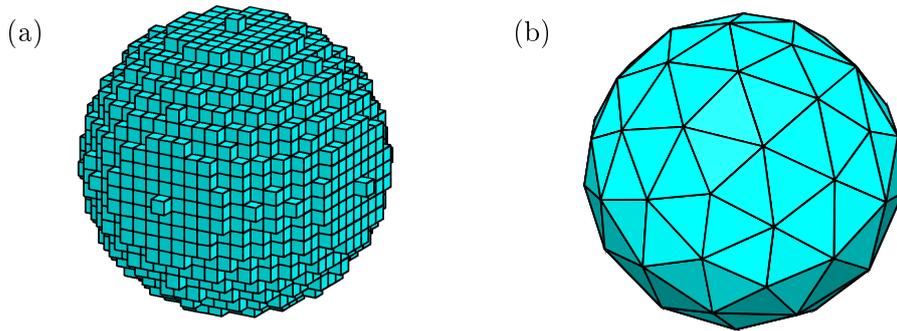
Figure 1.3: *A sphere discretized by (a) a rectangular grid and (b) an unstructured tetrahedral mesh. The sphere shown in (a) consists of 4169 cubes, while the mesh depicted in (b) contains only 118 tetrahedra.*

mous amounts of memory to still obtain converged results. To overcome these problems, it would be desirable to have a more sophisticated time-domain solver which employs adaptive meshes and allows for higher-order discretizations. Classical FEMs, as used in the frequency-domain, might seem an ideal choice. Unfortunately, they require the solution of a large system of equations at each timestep. The additional computational burden thereby renders traditional finite-element schemes unsuitable for time-domain calculations of larger nanophotonic systems. The problem can be circumvented by employing a variation of the FEM, namely a discontinuous Galerkin (DG) approach. This method was initially proposed by Hill and Reed [18] in the context of neutron transport. In the past few years, the DG method has attracted considerable attention and is now being employed to a wide range of problems from acoustics to hydrodynamics. Since the year 2002 there exists a mathematical proof for the convergence of the DG method when applied to Maxwell's equations [19].

In this thesis, we will carefully explore the applicability of both the FDTD and the DG method with respect to nanophotonic systems. After a brief reminder on classical electromagnetism (Ch. 2), we will use Ch. 3 to introduce a number of valuable analytical tools which enable us to model realistic nanophotonic setups. Next, in Ch. 4, the details of the FDTD method will be presented and its properties are thoroughly analyzed. In Ch. 5, we then proceed to an extensive discussion of the DG scheme. Here, we will already conduct first comparisons between the two methods on simple model systems. Before we continue with more realistic setups, we use Ch. 6 for a detour to the topic of time integration. In Ch. 7, we then come to the simulation of dielectric resonator structures. Next, we pursue a number of studies on metallic nanostructures in Ch. 8. There, we will also compare some of our results to experimental measurements. Finally, the thesis is concluded with a summary and an outlook in Ch. 9.

2 Chapter 2
Classical Electromagnetism

Before we can dive into the numerical simulation of nanophotonic systems, it makes sense to briefly recapitulate Maxwell's equations and some of their inherent properties. In the interest of brevity, we restrict the discussion to features relevant for the applications discussed in the later chapters. An exhaustive introduction to linear electromagnetism can be found in textbooks such as the one by David Jackson [20].

2.1 Maxwell's Equations

As discovered by J. C. Maxwell, the propagation of light is well described by the equations for electromagnetic radiation. Today, those equations are commonly called Maxwell's equations and in their differential formulation they read

$$\boxed{\begin{aligned}
\nabla \cdot \vec{D}(\vec{r}, t) &= \rho(\vec{r}, t), & \text{(2.1a)} \\
\nabla \cdot \vec{B}(\vec{r}, t) &= 0, & \text{(2.1b)} \\
\nabla \times \vec{E}(\vec{r}, t) &= -\frac{\partial \vec{B}(\vec{r}, t)}{\partial t}, & \text{(2.1c)} \\
\nabla \times \vec{H}(\vec{r}, t) &= \frac{\partial \vec{D}(\vec{r}, t)}{\partial t} + \vec{j}(\vec{r}, t). & \text{(2.1d)}
\end{aligned}}$$

Here, $\vec{E}(\vec{r}, t)$ and $\vec{H}(\vec{r}, t)$ denote the electric and magnetic field, respectively. Furthermore, $\vec{D}(\vec{r}, t)$ is called the electric displacement, while $\vec{B}(\vec{r}, t)$ signifies the magnetic induction. The charge density is described by $\rho(\vec{r}, t)$ and $\vec{j}(\vec{r}, t)$ denotes the electric currents in the system. To simplify the notation we subsequently drop the explicit temporal and spatial dependence of the variables unless it is imperative for the understanding.

By differentiating Eq. (2.1a) with respect to time and inserting Eq. (2.1d), we can derive the continuity equation

$$\frac{\partial}{\partial t}\rho + \nabla \cdot \vec{j} = 0, \tag{2.2}$$

which essentially states the conservation of charge in the system.

One can classify Maxwell's equations by the type of derivatives they contain. Thus, Eq. (2.1c) and (2.1d) are called the curl-equations, while Eq. (2.1a) and (2.1b) are termed divergence conditions. For the time-evolution only the curl-equations are relevant, while the divergence conditions can be seen as constraints which need to be fulfilled at all times. However, by applying the divergence to the curl-equations and using the continuity equation (2.2), it becomes obvious that the divergence of $\vec{D}$ and $\vec{B}$ is constant in time. Therefore, as long as the initial conditions fulfill the divergence conditions and the currents and charges obey the continuity equation, the curl-equations are sufficient to describe the entire time evolution of the electric and magnetic field.

In the Frequency Domain

In some cases it is advantageous to employ a Fourier transform and express Maxwell's equations in the frequency domain. We do so by using the convention

$$\frac{\partial}{\partial t} \overset{FT}{\longleftrightarrow} -i\omega$$

which yields

$$\nabla \cdot \breve{\vec{D}}(\vec{r}, \omega) = \breve{\rho}(\vec{r}, \omega), \tag{2.3a}$$

$$\nabla \cdot \breve{\vec{B}}(\vec{r}, \omega) = 0, \tag{2.3b}$$

$$\nabla \times \breve{\vec{E}}(\vec{r}, \omega) = i\omega \breve{\vec{B}}(\vec{r}, \omega), \tag{2.3c}$$

$$\nabla \times \breve{\vec{H}}(\vec{r}, \omega) = -i\omega \breve{\vec{D}}(\vec{r}, \omega) + \breve{\vec{j}}(\vec{r}, \omega). \tag{2.3d}$$

Here and for the rest of this thesis, $\breve{\ }$ denotes variables in the frequency domain and we will drop the explicit dependence on ω if not required for clarity.

Conservative Form

In order to obtain a more general formulation, it is possible to combine the curl-equations into a so-called conservative form

$$\partial_t \mathbf{u}(\vec{r}, t) + \partial_x \mathbf{F}_x(\mathbf{u}) + \partial_y \mathbf{F}_y(\mathbf{u}) + \partial_z \mathbf{F}_z(\mathbf{u}) = \mathbf{J}(\vec{r}, t). \tag{2.4}$$

Here, we have introduced $\mathbf{u} = \left(\vec{D}, \vec{B}\right)^T$, $\mathbf{J} = \left(\vec{j}, 0\right)^T$ and the flux vector

$$\mathbf{F}_i(\mathbf{u}) = \begin{pmatrix} -\hat{e}_i \times \vec{H} \\ \hat{e}_i \times \vec{E} \end{pmatrix}.$$

In this notation, $\hat{e}_i$, $i = x, y, z$ denote the Cartesian unit vectors. In order to explicitly see the conservative nature of Eq. (2.4), one can integrate them over a volume V with

the closed boundary ∂V. Using Gauss's law and neglecting the current terms then gives the expected form

$$\frac{\partial}{\partial t} \int_V \mathbf{u}(\vec{r}, t)\, d\vec{r} + \oint_{\partial V} \vec{F}(\mathbf{u}) \cdot \hat{n}\, da = 0,$$

where we have introduced $\vec{F} = \left(\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3\right)^T$ and the normal vector $\hat{n}$ of the boundary.

2.2 Constitutive Relations and Material Interfaces

So far, Maxwell's equations (2.1) contain twelve unknown field components with only eight independent equations. What is still missing are the constitutive relations, describing the interaction of the fields with matter. Using SI-units, these relations can be written as

$$\vec{B}(\vec{r}, t) = \mu_0 \vec{H}(\vec{r}, t) + \vec{M}(\vec{r}, t), \tag{2.5a}$$

$$\vec{D}(\vec{r}, t) = \epsilon_0 \vec{E}(\vec{r}, t) + \vec{P}(\vec{r}, t), \tag{2.5b}$$

where $\vec{P}$ and $\vec{M}$ are the macroscopic polarization and magnetization, respectively. The free-space permittivity ϵ_0 and the permeability μ_0 are natural constants which are related to the velocity of light in vacuum c by

$$c^2 = \frac{1}{\epsilon_0 \mu_0}.$$

The polarization and magnetization can both depend on the position $\vec{r}$, the time t and the fields $\vec{E}$ and $\vec{H}$ in a complicated manner. Obtaining a general expression is an intricate quantum mechanical problem and basically intractable. Fortunately, for most materials and reasonable light intensities, a full treatment is not necessary and excellent approximations can be made. For simplicity, we start with considering only local response in linear and isotropic media, which allows us to write

$$\vec{D}(\vec{r}, t) = \epsilon_0 \vec{E}(\vec{r}, t) + \epsilon_0 \int_{-\infty}^{\infty} dt' \chi_E(\vec{r}, t') \vec{E}(\vec{r}, t - t'), \tag{2.6a}$$

$$\vec{B}(\vec{r}, t) = \mu_0 \vec{H}(\vec{r}, t) + \mu_0 \int_{-\infty}^{\infty} dt' \chi_M(\vec{r}, t') \vec{H}(\vec{r}, t - t'), \tag{2.6b}$$

with the linear electric and magnetic susceptibilities $\chi_E(\vec{r}, t')$ and $\chi_M(\vec{r}, t')$, respectively. Above equations are still somewhat cumbersome because of the convolution integrals. By applying a Fourier transform, one arrives at the more convenient form

$$\breve{\vec{D}}(\vec{r}, \omega) = \epsilon_0 \breve{\epsilon}(\vec{r}, \omega) \breve{\vec{E}}(\vec{r}, \omega), \tag{2.7a}$$

$$\breve{\vec{B}}(\vec{r}, \omega) = \mu_0 \breve{\mu}(\vec{r}, \omega) \breve{\vec{H}}(\vec{r}, \omega). \tag{2.7b}$$

We have introduced the relative permittivity $\breve{\epsilon}(\vec{r}, \omega) = 1 + \breve{\chi}_E(\vec{r}, \omega)$ and the relative permeability $\breve{\mu}(\vec{r}, \omega) = 1 + \breve{\chi}_H(\vec{r}, \omega)$. Combining the constitutive relations from Eq. (2.7) with the curl equations in the frequency domain (2.3) then gives

$$\nabla \times \breve{\vec{E}}(\vec{r}, \omega) = i\omega\mu_0\mu(\vec{r}, \omega)\breve{\vec{H}}(\vec{r}, \omega), \tag{2.8a}$$

$$\nabla \times \breve{\vec{H}}(\vec{r}, \omega) = -i\omega\epsilon_0\epsilon(\vec{r}, \omega)\breve{\vec{E}}(\vec{r}, \omega) + \breve{\vec{j}}(\vec{r}, t). \tag{2.8b}$$

For the special case that both ϵ and μ are frequency-independent, i.e., non-dispersive, one can easily transform Eq. (2.7) back to the time-domain and obtain

$$\vec{D}(\vec{r}, t) = \epsilon_0\epsilon(\vec{r})\vec{E}(\vec{r}, t),$$
$$\vec{B}(\vec{r}, t) = \mu_0\mu(\vec{r})\vec{H}(\vec{r}, t).$$

In this case, the curl equations in the time domain can be simplified to

$$\nabla \times \vec{E}(\vec{r}, t) = -\mu_0\mu(\vec{r})\frac{\partial \vec{H}(\vec{r}, t)}{\partial t}, \tag{2.10a}$$

$$\nabla \times \vec{H}(\vec{r}, t) = \epsilon_0\epsilon(\vec{r})\frac{\partial \vec{E}(\vec{r}, t)}{\partial t} + \vec{j}(\vec{r}, t). \tag{2.10b}$$

Equivalently, we can also rewrite the conservative form (2.4) as

$$\mathcal{Q}\partial_t\mathbf{q} + \mathcal{A}_x\partial_x\mathbf{q} + \mathcal{A}_y\partial_y\mathbf{q} + \mathcal{A}_z\partial_z\mathbf{q} = 0, \tag{2.11}$$

where $\mathbf{q} = \left(\vec{E}, \vec{H}\right)^T$ and

$$\mathcal{Q} = \begin{pmatrix} \epsilon & 0 & 0 & 0 & 0 & 0 \\ 0 & \epsilon & 0 & 0 & 0 & 0 \\ 0 & 0 & \epsilon & 0 & 0 & 0 \\ 0 & 0 & 0 & \mu & 0 & 0 \\ 0 & 0 & 0 & 0 & \mu & 0 \\ 0 & 0 & 0 & 0 & 0 & \mu \end{pmatrix}, \quad \mathcal{A}_x = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\mathcal{A}_y = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \mathcal{A}_z = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Material Interfaces

All realistic setups will contain more than just one material, which means that we will have to deal with interfaces. The behavior of the electric and magnetic fields at such interfaces can be obtained by using Stokes's and Gauss's theorem [20]. Assuming

an interface with normal vector $\hat{n}$, pointing from medium 1 to medium 2. When integrating the curl equations over a surface across the interface and applying Stokes's theorem we obtain the conditions

$$\hat{n} \times \left(\vec{E}_2 - \vec{E}_1 \right) = 0 \quad \text{and} \quad \hat{n} \times \left(\vec{H}_2 - \vec{H}_1 \right) = \vec{j}_s,$$

where $\vec{j}_s$ denotes the surface current density. For vanishing $\vec{j}_s$ we see that the tangential components of $\vec{E}$ and $\vec{H}$ are continuous across the interface. Similarly, integrating the divergence conditions over an infinitesimal volume across the interface and applying Gauss's theorem yields

$$\hat{n} \cdot \left(\vec{D}_2 - \vec{D}_1 \right) = \rho_s \quad \text{and} \quad \hat{n} \cdot \left(\vec{B}_2 - \vec{B}_1 \right) = 0.$$

Here, ρ_s signifies the surface charge density. Neglecting ρ_s, we observe that the normal components of $\vec{D}$ and $\vec{B}$ are continuous. As a consequence, the normal components of $\vec{E}$ and $\vec{H}$ are discontinuous and for linear media the magnitude of the jumps is given by

> **Boundary Conditions**
>
> $$\frac{\hat{n} \cdot \vec{E}_2}{\hat{n} \cdot \vec{E}_1} = \frac{\epsilon_1}{\epsilon_2} \quad \text{and} \quad \frac{\hat{n} \cdot \vec{H}_2}{\hat{n} \cdot \vec{H}_1} = \frac{\mu_1}{\mu_2}.$$
>
> $\qquad(2.12)$

As we will demonstrate in the later chapters, it is precisely these discontinuities that complicate the numerical treatment of Maxwell's equations.

2.3 Symmetries and Reduction to Lower Dimensions

In some cases, it is not necessary to solve the full set of Maxwell's equations, due to symmetries in the underlying physical system. In practice, relevant symmetries are either translational or rotational ones. Here, we restrict ourselves to translational symmetries, which lead to an effective reduction of the dimensionality of the system.

Reduction to Two Dimensions

If the entire system is homogeneous in z-direction, this corresponds to an effectively two-dimensional system. Then, all z-derivatives vanish and instead of the six coupled equations, we obtain two decoupled sets of three equations. Explicitly, those sets read

> **TE Polarization**
>
> $$\partial_t E_x = \frac{1}{\epsilon_0 \epsilon} \left(\partial_y H_z - j_x \right),$$
>
> $\qquad(2.13a)$
>
> $$\partial_t E_y = -\frac{1}{\epsilon_0 \epsilon} \left(\partial_x H_z - j_y \right),$$
>
> $\qquad(2.13b)$
>
> $$\partial_t H_z = \frac{1}{\mu_0 \mu} \left(\partial_y E_x - \partial_x E_y \right),$$
>
> $\qquad(2.13c)$

and

<div style="border:1px solid; background:#fbf8cc; padding:10px;">

TM Polarization

$$\partial_t H_x = -\frac{1}{\mu_0 \mu} \partial_y E_z, \tag{2.14a}$$

$$\partial_t H_y = \quad \frac{1}{\mu_0 \mu} \partial_x E_z, \tag{2.14b}$$

$$\partial_t E_z = \quad \frac{1}{\epsilon_0 \epsilon} \left(\partial_x H_y - \partial_y H_x - j_z \right). \tag{2.14c}$$

</div>

Equations (2.13) describe propagation in the so-called transverse-electric (TE) polarization, where the electrical field vector lies in the (x,y)-plane. In contrast, the transverse-magnetic (TM) polarization as described by Eqs. (2.14) possesses an electrical field perpendicular to the plane of propagation[1].

Reduction To One Dimension

If there is a further translational symmetry along the y-direction, it is possible to reduce Eq. (2.13) and Eq. (2.14) to a single set of two equations,

<div style="border:1px solid; background:#fbf8cc; padding:10px;">

One-Dimensional System

$$\partial_t H_y = \frac{1}{\mu_0 \mu} \partial_x E_z, \tag{2.15a}$$

$$\partial_t E_z = \frac{1}{\epsilon_0 \epsilon} \left(\partial_x H_y - j_z \right). \tag{2.15b}$$

</div>

In this case, we usually drop the indices of the field components.

2.4 Plane Wave Solutions

As alluded to previously, there are no known analytical solutions to Maxwell's equations with general material distributions. However, for the particularly simple case of a homogeneous medium with spatially constant ϵ and μ, general solutions can be readily found. To simplify the derivation, we start from the curl equations in the frequency domain (2.8). Solving Eq. (2.8a) for $\breve{\vec{H}}$ and inserting it into Eq. (2.8b) gives

$$\nabla \times \nabla \times \breve{\vec{E}} = \frac{\omega^2}{c^2} \breve{\epsilon} \breve{\mu} \breve{\vec{E}} + i\omega \mu_0 \breve{\mu} \breve{\vec{j}}.$$

---

[1]Unfortunately, the nomenclature of "TE" and "TM" is not unequivocal in the literature. Especially in the context of planar waveguides, the terms are used with the exact opposite meaning. Therefore care needs to be taken, when discussing the polarization of light in a two-dimensional system.

Exploiting the vector identity $\nabla \times \nabla \times \breve{\vec{E}} = \nabla \cdot \left(\nabla \cdot \breve{\vec{E}} \right) - \nabla^2 \cdot \breve{\vec{E}}$ and inserting the divergence condition (2.1a) one obtains the wave equation

$$
\boxed{\text{The Wave Equation} \atop \nabla^2 \cdot \breve{\vec{E}} - \frac{\omega^2}{c^2} \breve{n}^2 \breve{\vec{E}} = i \omega \mu_0 \breve{\mu} \breve{\vec{j}}.}
$$
(2.16)

Here, we introduced the index of refraction $\breve{n}^2 = \breve{\mu} \breve{\epsilon}$. If we assume vanishing currents, solutions of Eq. (2.16) can be found by making a plane wave ansatz

$$
\vec{E}(\vec{r}, t) = \vec{E_0} e^{i(\vec{k}\vec{r} - \omega t)}
$$
(2.17)

where $\vec{E_0}$ is the polarization of the wave, $\vec{k}$ is the wave vector and ω is the corresponding angular frequency. In the frequency domain, the expression reads

$$
\breve{\vec{E}}(\vec{r}, \omega') = \vec{E_0} e^{i\vec{k}\vec{r}} \delta(\omega - \omega'),
$$

where δ denotes the Dirac δ-function. The wave vector and the frequency are connected via the dispersion relation

$$
\omega^2 = \frac{c^2}{\breve{n}^2(\omega)} \left| \vec{k} \right|^2.
$$

By inserting the ansatz (2.17) into Eq. (2.1a), one directly obtains the property that $\vec{E_0} \perp \vec{k}$. From Eq. (2.1d) we find the magnetic field to be

$$
\vec{H} = \frac{1}{\mu_0 \mu \omega} \vec{k} \times \vec{E} = Z(\omega)^{-1} \frac{\vec{k}}{\left| \vec{k} \right|} \times \vec{E},
$$
(2.18)

where we have introduced the impedance $Z(\omega) = \sqrt{\frac{\mu_0 \mu(\omega)}{\epsilon_0 \epsilon(\omega)}}$. From this result we observe that $\vec{E}$, $\vec{H}$ and $\vec{k}$ form an orthogonal trihedral, so

$$
\vec{E} \perp \vec{k}, \quad \vec{H} \perp \vec{k}, \quad \vec{E} \perp \vec{H}.
$$

Reflection and Transmission at a Plane Interface

With the plane wave solutions for a homogeneous medium, we can ask the question of what happens at a plane interface between two homogeneous half spaces. Such a setup is sketched in Fig. 2.1, where we assume an incoming wave with frequency ω and wave vector $\vec{k_I}$. The interface leads to a splitting of the wave into a reflected and a transmitted part. The full ansatz reads

$$
\vec{E} = \begin{cases} \vec{E_I} + \vec{E_R} & \text{if } x < 0 \\ \vec{E_T} & \text{if } x > 0 \end{cases} \quad \text{and} \quad \vec{H} = \begin{cases} \vec{H_I} + \vec{H_R} & \text{if } x < 0 \\ \vec{H_T} & \text{if } x > 0 \end{cases},
$$
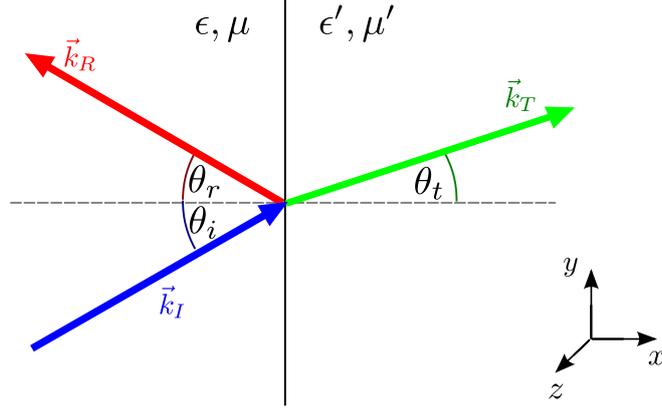
Figure 2.1: *Sketch of an electromagnetic wave with wave vector $\vec{k}_I$ impinging on an interface. The wave is split into a reflected part with wave vector $\vec{k}_R$ and a transmitted part with $\vec{k}_T$.*

where

$$\vec{E}_I = \vec{E}_{I0}\, e^{i(\vec{k}_I \vec{r} - \omega t)}, \qquad\qquad \vec{H}_I = \frac{1}{\mu_0 \mu \omega} \vec{k}_I \times \vec{E}_I,$$

$$\vec{E}_R = \vec{E}_{R0}\, e^{i(\vec{k}_R \vec{r} - \omega t)}, \qquad\qquad \vec{H}_R = \frac{1}{\mu_0 \mu \omega} \vec{k}_R \times \vec{E}_R,$$

$$\vec{E}_T = \vec{E}_{T0}\, e^{i(\vec{k}_T \vec{r} - \omega t)}, \qquad\qquad \vec{H}_T = \frac{1}{\mu_0 \mu' \omega} \vec{k}_T \times \vec{E}_T.$$

Here, we have introduced

$$k = |\vec{k}_I| = |\vec{k}_R| = \frac{\omega}{c} n \quad \text{and} \quad k' = |\vec{k}_T| = \frac{\omega}{c} n'.$$

Since the boundary conditions (2.12) are independent of time and must be fulfilled at all points on the interface, we conclude

$$\vec{k}_I \vec{r}\Big|_{x=0} = \vec{k}_R \vec{r}\Big|_{x=0} = \vec{k}_T \vec{r}\Big|_{x=0}.$$

As a consequence, all the $\vec{k}$-vectors must lie in a plane (plane of incidence) and we can choose our coordinate system such that the z-component of all $\vec{k}$-vectors vanishes. Furthermore, we see that all y-components of the $\vec{k}$-vectors are equal and the x-components of $\vec{k}_I$ and $\vec{k}_R$ must have equal length but opposite sign

$$k_y = k_{Iy} = k_{Ry} = k_{Ty} \quad \text{and} \quad k_{Ix} = -k_{Rx}.$$

In order to proceed, it makes sense to distinguish two cases:

- $\vec{E}$ perpendicular to the plane of incidence (s-polarization)

- $\vec{E}$ parallel to the plane of incidence (p-polarization)

We will start with the first case (s-polarization), which allows us to write

$$\vec{E}_{I0} = \hat{e}_z, \ \vec{E}_{R0} = R_s \hat{e}_z, \ \vec{E}_{T0} = T_s \hat{e}_z,$$

and reduces the magnetic fields to

$$\vec{H}_I = \frac{1}{\mu_0 \mu \omega} \left[k_{Iy} \hat{e}_x - k_{Ix} \hat{e}_y \right] e^{i(\vec{k}_I \vec{r} - \omega t)},$$

$$\vec{H}_R = R_s \frac{1}{\mu_0 \mu \omega} \left[k_{Ry} \hat{e}_x - k_{Rx} \hat{e}_y \right] e^{i(\vec{k}_R \vec{r} - \omega t)},$$

$$\vec{H}_T = T_s \frac{1}{\mu_0 \mu' \omega} \left[k_{Ty} \hat{e}_x - k_{Tx} \hat{e}_y \right] e^{i(\vec{k}_T \vec{r} - \omega t)}.$$

Next, we apply the boundary conditions (2.12) to our fields. The continuity of the tangential components results in

$$1 + R_s = T_s \quad \text{and} \quad \frac{1}{\mu} k_{Ix} \left(1 - R_s \right) = \frac{1}{\mu'} k_{Tx} T_s.$$

Combining the two expressions, we obtain the reflection amplitude R_s as

$$R_s = \frac{k_{Ix}/\mu - k_{Tx}/\mu'}{k_{Ix}/\mu + k_{Tx}/\mu'} = \frac{\sqrt{k^2 - k_y^2}/\mu - \sqrt{k'^2 - k_y^2}/\mu'}{\sqrt{k^2 - k_y^2}/\mu + \sqrt{k'^2 - k_y^2}/\mu'}.$$

By using the incident angle θ_i as defined in Fig. 2.1 we obtain the final results for the reflected and transmitted amplitudes as

$$R_s = \frac{n \cos(\theta_i)/\mu - \sqrt{n'^2 - n^2 \sin^2(\theta_i)}/\mu'}{n \cos(\theta_i)/\mu + \sqrt{n'^2 - n^2 \sin^2(\theta_i)}/\mu'}, \tag{2.19a}$$

$$T_s = \frac{2n \cos(\theta_i)/\mu}{n \cos(\theta_i)/\mu + \sqrt{n'^2 - n^2 \sin^2(\theta_i)}/\mu'}. \tag{2.19b}$$

Those equations are commonly known as Fresnel's equations for s-polarization. The equivalent formulae for p-polarization can be obtained easily by exploiting a symmetry of Maxwell's equation [20]. Namely, we can replace $\vec{E} \to \vec{H}$, $\vec{H} \to -\vec{E}$ and $\epsilon \leftrightarrow \mu$. The sign-change does not contribute in the above calculations, so it suffices to swap ϵ and μ to obtain

$$R_p = \frac{n \cos(\theta_i)/\epsilon - \sqrt{n'^2 - n^2 \sin^2(\theta_i)}/\epsilon'}{n \cos(\theta_i)/\epsilon + \sqrt{n'^2 - n^2 \sin^2(\theta_i)}/\epsilon'}, \tag{2.20a}$$

$$T_p = \frac{2n \cos(\theta_i)/\epsilon}{n \cos(\theta_i)/\epsilon + \sqrt{n'^2 - n^2 \sin^2(\theta_i)}/\epsilon'}. \tag{2.20b}$$

It should be noted that Fresnel's equations are equally valid for dispersive media, but they do rely on the isotropy of the system. We will return to this point in chapter 3.2.

2.5 Dyadic Green's Tensor and Sources

While the plane wave solutions discussed in the previous section provide some insight into the propagation of light, they are not very helpful in practice. The problem is that plane waves are infinitely extended in both time and space, while in reality one usually has to deal with pulses of finite extent. Such pulses are typically introduced by a current distribution $\vec{j}(\vec{r}, t) = \vec{J}(\vec{r})j(t)$, where we have separated the spatial distribution $\vec{J}(\vec{r})$ from its time dependence $j(t)$. For simplicity, we will further assume that the source starts to radiate at time $t = 0$ and all fields are zero before that time. Under these conditions, a solution of Maxwell's equations can be expressed in terms of the dyadic Green's tensors $\underline{\underline{G}}_E$ and $\underline{\underline{G}}_H$ as

$$\vec{E}(\vec{r}, t) = \int dt' \int d\vec{r}' \, \underline{\underline{G}}_E(\vec{r}, t | \vec{r}', t') \vec{J}(\vec{r}') j(t'), \tag{2.21a}$$

$$\vec{H}(\vec{r}, t) = \int dt' \int d\vec{r}' \, \underline{\underline{G}}_H(\vec{r}, t | \vec{r}', t') \vec{J}(\vec{r}') j(t'). \tag{2.21b}$$

The derivation of the Green's tensors is a very tedious task and a general procedure is described in Ref. [21]. The results strongly depend on the dimensionality of the system.

The One-Dimensional Case

For a one-dimensional system, the Green's tensors are scalar and read

$$\underline{\underline{G}}_E^{1D}(x, t | x', t') = \frac{\Theta(\tau)}{2} \delta(|r| - \tau), \tag{2.22a}$$

$$\underline{\underline{G}}_H^{1D}(x, t | x', t') = Z^{-1} \frac{\Theta(\tau)}{2} \operatorname{sgn}(r) \delta(|r| - \tau), \tag{2.22b}$$

where, for brevity, we introduced $\tau = \frac{c}{n}(t - t')$ and $r = x - x'$. Further, Θ denotes the Heaviside step function. In this case, one can directly observe how a source generates right- and left-moving fields. The temporal shape of the source is conserved and it symmetrically travels with the speed c/n in both directions.

The Two-Dimensional Case in TM-Polarization

For the two-dimensional case, we have to distinguish between the two possible polarizations TE and TM. Here, we only give the expressions for TM-polarization, where the vector $\vec{J}(\vec{r})$ is perpendicular to the plane of propagation. Then, the Green's tensors in polar coordinates take the form

$$\underline{\underline{G}}_E^{TM}(\rho, \varphi, \tau) = \frac{\Theta(\tau)}{2\pi} \left[\frac{\delta(\tau - \rho)}{\sqrt{\tau^2 - \rho^2}} - \frac{\tau \Theta(\tau - \rho)}{\left(\tau^2 - \rho^2\right)^{3/2}} \right] \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.23a}$$

and

$$\underline{\underline{G}}_H^{\mathrm{TM}}(\rho,\varphi,\tau) = \frac{\Theta(\tau)}{2\pi Z}\left[\frac{\delta(\tau-\rho)}{\sqrt{\tau^2-\rho^2}} - \frac{\rho\Theta(\tau-\rho)}{\left(\tau^2-\rho^2\right)^{3/2}}\right]\begin{pmatrix} 0 & 0 & -\cos(\varphi) \\ 0 & 0 & \sin(\varphi) \\ 0 & 0 & 0 \end{pmatrix}. \qquad (2.23b)$$

From these expression, we can already identify a significant problem, namely the singularities at $\rho = 0$ and $\rho = \tau$. In the next section, we will give an example of how to overcome those problems and how to obtain a numerical solution from Eqs. (2.23). The case of TE-polarization, where the source lies in-plane, is slightly more cumbersome but does not offer any new insight. The explicit expressions for the TE case can be derived from the tensors given in Ref. [21].

The Three-Dimensional Case

In the fully three dimensional case, both $\underline{\underline{G}}_E^{\mathrm{3D}}$ and $\underline{\underline{G}}_H^{\mathrm{3D}}$ become 3×3-matrices. Due to inherent symmetries, they can be written as

$$\underline{\underline{G}}_E^{\mathrm{3D}} = \begin{pmatrix} A_1 & B_3 & B_2 \\ B_3 & A_2 & B_1 \\ B_2 & B_1 & A_3 \end{pmatrix} \quad \text{and} \quad \underline{\underline{G}}_H^{\mathrm{3D}} = \begin{pmatrix} 0 & C_3 & C_2 \\ -C_3 & 0 & C_1 \\ -C_2 & -C_1 & 0 \end{pmatrix}, \qquad (2.24)$$

where

$$A_i = \left(S - 3L\right)\frac{d_i^2}{R^2} + L + T,$$

$$B_1 = \left(S - 3L\right)\frac{d_2\,d_3}{R^2}, \quad B_2 = \left(S - 3L\right)\frac{d_1\,d_3}{R^2}, \quad B_3 = \left(S - 3L\right)\frac{d_1\,d_2}{R^2},$$

$$C_i = (-1)^i\,Z^{-1}P\frac{d_i}{R}$$

with

$$L = \frac{\Theta(\tau)}{4\pi}\left[\frac{\delta(\tau-R) - \delta(R)}{R^2} + \frac{\Theta(\tau-R) - \Theta(-R)}{R^3}\right],$$

$$P = \frac{\Theta(\tau)}{4\pi}\left[\frac{\delta(\tau-R)}{R^2} + \frac{\delta'(\tau-R)}{R}\right],$$

$$S = -\frac{\Theta(\tau)}{4\pi}\frac{\delta'(\tau-R) + \delta'(R)}{R},$$

$$T = \frac{\Theta(\tau)}{4\pi}\frac{\delta'(\tau-R)}{R},$$

and

$$\vec{d} = \vec{r} - \vec{r}'. \qquad (2.25)$$

For a general spatial distribution and an arbitrary time dependence of $\vec{j}(\vec{r}, t)$, an explicit solution of the integrals in Eqs. (2.21) will not be feasible. In those cases, one has to resort to numerical integration, which also is highly problematic due to the step- and delta-functions occurring. However, as we will demonstrate in the following, for particular spatial distributions of $\vec{j}$, it is possible to evaluate Eqs. (2.21) by a combination of analytical and numerical methods.

2.5.1 Fields Generated by a Point Source

As a particularly important example, we consider the case where $\vec{j}(\vec{r}, t) = \vec{j}_0 \delta(\vec{r}) j(t)$. This corresponds to a point source at the origin with orientation $\vec{j}_0$ and time dependence $j(t)$.

The Two-Dimensional Case in TM Polarization

In the case of TM polarization, the orientation $\vec{j}_0$ is fixed along the z-axis. Inserting the Green's tensors from Eq. (2.23) into Eq. (2.21) and performing the spatial integration leaves us with

$$E_z(\rho, \varphi, t) = \frac{1}{2\pi} \left[\int_0^t \frac{\delta(\tau - \rho)}{\sqrt{\tau^2 - \rho^2}} j(t') \, dt' - \int_0^t \frac{\tau \Theta(\tau - \rho)}{(\tau^2 - \rho^2)^{3/2}} j(t') \, dt' \right].$$

In the next step, we change the integration variable from t' to τ and perform an integration by parts on the second integral. One of the occurring boundary terms will exactly cancel with the first integral and the remaining expression reads

$$E_z(\rho, \varphi, t) = \frac{1}{2\pi} \int_\rho^{\frac{c}{n}t} \frac{j'(\frac{c}{n}t - \tau)}{\sqrt{\tau^2 - \rho^2}} \, d\tau = \frac{1}{2\pi} \int_\rho^{\frac{c}{n}t} \frac{j'(\frac{c}{n}t - \tau)}{\sqrt{\tau - \rho}\sqrt{\tau + \rho}} \, d\tau.$$

Here, j' denotes the derivative of the source term with respect to its argument. Unfortunately, the integral contains a singularity at the lower limit $\tau = \rho$. However, as can be seen from the second expression, the singularity is integrable. By using a tailored quadrature [22] one can obtain accurate numerical results for arbitrary time envelopes. The procedure for the two magnetic field components is almost identical and yields

$$\begin{pmatrix} H_x(\rho, \varphi, t) \\ H_y(\rho, \varphi, t) \end{pmatrix} = \frac{1}{2\pi Z} \int_\rho^{\frac{c}{n}t} \frac{\tau j'(\frac{c}{n}t - \tau)}{\sqrt{\tau - \rho}\sqrt{\tau + \rho}} \, d\tau \begin{pmatrix} -\cos(\varphi) \\ \sin(\varphi) \end{pmatrix}.$$

The Three-Dimensional Case

Essentially, the three-dimensional cases executes as the previous one, except for the more cumbersome expressions. The final result can be stated explicitly as

$$\vec{E}(\vec{r}, t) = \left(\frac{\bar{j}(t)}{4\pi R^5} + \frac{j(t)}{4\pi R^4} \right) \left(\vec{J} - 3\vec{r} \left(\vec{r} \cdot \vec{J} \right) \right) + \frac{j'(t)}{4\pi R^3} \left(\vec{J} - \vec{r} \left(\vec{r} \cdot \vec{J} \right) \right),$$

$$\vec{H}(\vec{r}, t) = \left(\frac{j(t)}{4\pi R^5} + \frac{j'(t)}{4\pi R^4} \right) \left(\vec{r} \times \vec{J} \right),$$

where we have introduced the integral

$$\bar{j}(t) = \int_{t_0}^{t} j(t') dt'.$$

It should be noted that, in contrast to the two-dimensional case, there are no singularities or divergences to overcome, as long as $j(t)$ is a well-behaved function. For typical time-dependencies like Gaussian pulses or sinusoidal oscillations, the integral in $\bar{j}(t)$ can even be evaluated analytically, resulting in an entirely explicit expression for the fields.

2.6 Energy Density and the Poynting Vector

Experimentally, the light intensity, which corresponds to the energy flux of the electromagnetic field, plays an important role. To find an explicit expression for the energy flux, we start from the curl equations (2.1c) and (2.1d), which can be combined to

$$\vec{E} \cdot \left(\nabla \times \vec{H} \right) - \vec{H} \cdot \left(\nabla \times \vec{E} \right) = \vec{j} \cdot \vec{E} + \vec{E} \cdot \frac{\partial}{\partial t} \vec{D} + \vec{H} \cdot \frac{\partial}{\partial t} \vec{B}. \tag{2.26}$$

Using the vector identity $\nabla \cdot \left(\vec{E} \times \vec{H} \right) = \vec{H} \cdot \left(\nabla \times \vec{E} \right) - \vec{E} \cdot \left(\nabla \times \vec{H} \right)$, we can transform Eq. (2.26) to

$$\vec{E} \cdot \frac{\partial}{\partial t} \vec{D} + \vec{H} \cdot \frac{\partial}{\partial t} \vec{B} + \nabla \cdot \left(\vec{E} \times \vec{H} \right) = -\vec{j} \cdot \vec{E}. \tag{2.27}$$

For linear and isotropic media, we can then use the relation

$$\vec{E} \cdot \frac{\partial}{\partial t} \vec{D} + \vec{H} \cdot \frac{\partial}{\partial t} \vec{B} = \frac{\partial}{\partial t} \left(\frac{1}{2} \vec{E} \cdot \vec{D} + \frac{1}{2} \vec{H} \cdot \vec{B} \right)$$

to reformulate Eq. (2.27) as

$$\frac{\partial}{\partial t} \underbrace{\left(\frac{1}{2} \vec{E} \cdot \vec{D} + \frac{1}{2} \vec{H} \cdot \vec{B} \right)}_{w} + \nabla \cdot \underbrace{\left(\vec{E} \times \vec{H} \right)}_{\vec{S}} = -\vec{j} \cdot \vec{E}. \tag{2.28}$$

This expression is known as Poynting's theorem. Calculating the work done by the fields on a charged body [23] allows us to identify w with the electromagnetic energy density. For vanishing currents $\vec{j}$, we also recover that Eq. (2.28) has the form of a conservation law. Therefore, we can interpret the vector $\vec{S}$ as the energy flux density. This quantity is also called the Poynting vector.

In the case of time-harmonic fields $\breve{\vec{E}}$ and $\breve{\vec{H}}$, we are typically interested in the time-averaged energy flux as a function of frequency. The proper expression in this case is given by

$$\vec{S} = \frac{1}{2}\mathrm{Re}\left[\breve{\vec{E}} \times \breve{\vec{H}}^*\right], \tag{2.29}$$

where the $*$ denotes the complex conjugate. It should be noted that the energy flux in the frequency domain can *not* be obtained by the direct Fourier transform of $\vec{S}(\vec{r}, t)$. As an argument for why this would be wrong, one should consider a single plane wave with frequency ω_0, i.e., $\vec{E}(\vec{r}, t) = \mathrm{Re}\left[\vec{E}_0 e^{-i(\vec{k}\vec{r} - \omega_0 t)}\right]$. According to Eq. (2.18), the magnetic field then takes the form $\vec{H}(\vec{r}, t) = Z\frac{\vec{k}}{|\vec{k}|} \times \vec{E}$. Inserting this into the expression for the Poynting vector gives

$$\vec{S} = Z\left(\vec{E} \times \frac{\vec{k}}{|\vec{k}|} \times \vec{E}\right) = Z|\vec{E}_0|^2 \cos^2\left(\vec{k}\vec{r} - \omega_0 t\right)\frac{\vec{k}}{|\vec{k}|}.$$

Using the relation $\cos^2(x) = \left[1 + \cos(2x)\right]/2$ and applying a Fourier transform would lead to a spectrum with delta-peaks at 0 and $2\omega_0$. This can clearly not be the correct result for a plane wave with carrier frequency ω_0.

2.7 Rescaling to Dimensionless Units

So far, we have worked with Maxwell's equations in SI units, which introduced the constants ϵ_0, μ_0 and c. However, an interesting feature of Maxwell's equations in non-dispersive media is, that they do not contain a fundamental length scale. Furthermore, if the media are linear, we are also free to scale the field strengths arbitrarily. Consequently, it is possible to rescale Maxwell's equations, making them independent of any unit system.

To this end, we introduce an arbitrary length scale a and an arbitrary field strength E_0 to rescale the individual variables as follows:

$$\vec{r} \to \tilde{\vec{r}}a, \quad t \to \tilde{t}a/c, \quad \vec{E} \to \tilde{\vec{E}}E_0, \quad \vec{H} \to \tilde{\vec{H}}\sqrt{\frac{\epsilon_0}{\mu_0}}E_0, \quad \vec{j} \to \tilde{\vec{j}}\sqrt{\frac{\epsilon_0}{\mu_0}}\frac{E_0}{a}.$$

Using these transformations yields the rescaled curl equations

$$\partial_{\tilde{t}}\tilde{\vec{E}} = \frac{1}{\epsilon}\left(\tilde{\nabla} \times \tilde{\vec{H}} - \tilde{\vec{j}}\right), \tag{2.30a}$$

$$\partial_{\tilde{t}}\tilde{\vec{H}} = -\frac{1}{\mu}\tilde{\nabla} \times \tilde{\vec{E}}, \tag{2.30b}$$

	SI→Dimensionless	Dimensionless→SI
Length	$\tilde{l} = l\frac{1}{a}$	$l = \tilde{l}a$
Time	$\tilde{t} = t\frac{c}{a}$	$t = \tilde{t}\frac{a}{c}$
Frequency	$\tilde{\nu} = \nu\frac{a}{c}$	$\nu = \tilde{\nu}\frac{c}{a}$
Electric Field	$\tilde{\vec{E}} = \vec{E}/E_0$	$\vec{E} = \tilde{\vec{E}}E_0$
Magnetic Field	$\tilde{\vec{H}} = \vec{H}/\left(\sqrt{\frac{\epsilon_0}{\mu_0}}E_0\right)$	$\vec{H} = \tilde{\vec{H}}\left(\sqrt{\frac{\epsilon_0}{\mu_0}}E_0\right)$

Table 2.1: *Table with conversion factors between quantities in SI units (without tilde) and quantities in dimensionless units (with tilde).*

where all variables are now dimensionless. Thus, the velocity of light in vacuum is now unity. Subsequently, we will drop the tilde and always work in dimensionless units, unless stated otherwise. For a given length scale a and a chosen field strength E_0 one can translate variables from dimensionless units into SI units and vice versa by using the factors given in Tab. 2.1. For the rest of this thesis we will work with dimensionless units, unless stated otherwise.

3

Auxiliary Differential Equations

Before we can proceed to the actual numerical methods, there are a few issues that deserve our attention. As discussed in the previous chapter, material dispersion leads to convolution integrals in the constitutive relations. Thus, Maxwell's equations become a set of integro-differential equations, which significantly complicates numerical treatment. A second, seemingly independent problem is the truncation of the computational domain. In particular, we need a way to absorb all outgoing radiation without reflections. This is achieved by a thin layer of a specially tailored material, called a perfectly matched layer (PML). As we will see, such an absorbing material necessarily is dispersive.

In this chapter, we will modify Maxwell's equations in order to include both dispersive media and PMLs without convolution integrals. The key to the implementation lies in a technique called auxiliary differential equations (ADEs). This analytical method allows for certain models of dispersion to transform the convolution integrals into a set of differential equations. Those additional equations can then be solved in parallel with Maxwell's equations.

3.1 Models of Dispersion

In nature, essentially all optical media show some form of material dispersion, i.e., their refractive index depends on the frequency of light. For most glasses, this dispersion is rather weak in the visible range and only plays a major role if we want to work in a broad spectral range. Metals on the other hand are strongly dispersive and one must always account for their frequency dependent properties. As we have seen in Sec. 2.2, dispersion leads to convolution integrals (2.6) in the time-domain. Therefore, we are not directly able to include arbitrary dielectric functions when working with a time-domain formulation. Fortunately, this is not necessary since many of the naturally occurring permittivities can be approximated by rather simple analytical models. The idea of these models is to describe the microscopic response of the material via an induced microscopic dipole moment $\vec{p}$. For a given $\vec{p}$, one can obtain the macroscopic
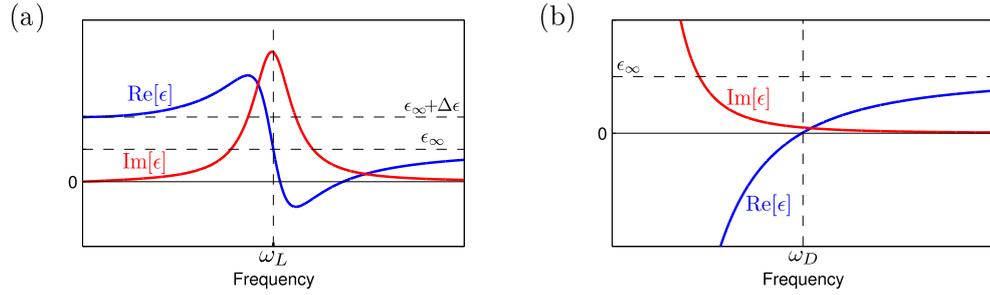
Figure 3.1: *A sketch of the dispersion relations: (a) Lorentz model, (b) Drude model.*

polarization $\vec{P}$ by

$$\vec{P} = n\vec{p}, \tag{3.1}$$

where n is the density of microscopic dipole moments.

3.1.1 The Lorentz Model

As a first model, we assume the electrons in the material to be described as classical harmonic oscillators that are driven by the incoming electric field. The equation of motion for the position $\vec{x}$ of an individual electron is then given by

$$\frac{\partial^2}{\partial t^2}\vec{x} + \gamma_L \frac{\partial}{\partial t}\vec{x} + \omega_L^2 \vec{x} = -\frac{e}{m_e}\vec{E}.$$

Here, we introduced the eigenfrequency ω_L, a dissipation term γ_L and the coupling constant $\frac{e}{m_e}$ to the electric field. Applying a Fourier transform, we can directly solve this equation for the position

$$\breve{\vec{x}}(\omega) = -\frac{1}{\omega_L^2 - i\gamma_L\omega - \omega^2}\frac{e}{m_e}\breve{\vec{E}}.$$

Using the definition of a dipole moment as $\breve{\vec{p}} = e\breve{\vec{x}}$ and inserting it into Eq. (3.1), we obtain the dielectric function from $\breve{\vec{D}} = \epsilon_\infty \breve{\vec{E}} + \breve{\vec{P}} = \breve{\epsilon}\breve{\vec{E}}$ as

$$\boxed{\begin{array}{c} \text{Lorentz Model} \\ \breve{\epsilon}(\omega) = \epsilon_\infty + \frac{\Delta\epsilon\,\omega_L^2}{\omega_L^2 - i\gamma_L\omega - \omega^2}. \end{array}} \tag{3.2}$$

We have defined $\Delta\epsilon = \frac{ne^2}{\epsilon_0 m_e \omega_L^2}$ and we also generalized the model slightly by introducing ϵ_∞. Eq. (3.2) is commonly called Lorentz model and the shape of the dielectric function is plotted in Fig. 3.1(a).

It should be noted that the Lorentz model does not serve well as an ab-initio description of matter. Instead, it can be used to fit the experimentally measured dielectric

function in a certain frequency range. A sum of several Lorentz terms without absorption ($\gamma_L = 0$) is used in optics to describe the dielectric function of different glasses. This description is called Sellmeier formula and reads

$$\epsilon(\lambda) = A + \sum_i \frac{B_i \lambda^2}{\lambda^2 - C_i^2},$$

where λ is the wavelength and A, B_i, C_i are material parameters. For common optical glasses, using this formula with $i = 3$ terms can reproduce the experimental data in the visible spectrum with a relative error below 10^{-5} [24].

3.1.2 The Drude Model

For the derivation of the Lorentz model, we assumed localized electrons which are bound to their nuclei. For metals, this description fails and it makes more sense to assume that some electrons can move freely within the material. In contrast to the Lorentz model, we now set up an equation for the velocity $\vec{v}$ of a single electron as

$$\frac{\partial}{\partial t}\vec{v} + \gamma_D \vec{v} = -\frac{e}{m_e}\vec{E}. \tag{3.3}$$

Executing a Fourier transform then gives

$$\breve{\vec{v}} = \frac{1}{i\omega - \gamma_D}\frac{e}{m_e}\breve{\vec{E}}, \tag{3.4}$$

from which we can derive a current density $\breve{\vec{j}}$ as

$$\breve{\vec{j}} = -en\breve{\vec{v}} = -\frac{ne^2/m_e}{i\omega - \gamma_D}\breve{\vec{E}}. \tag{3.5}$$

Inserting this current into the wave equation (2.16) with a dispersionless index of refraction then yields

$$\nabla^2 \cdot \breve{\vec{E}} - \omega^2 \mu \underbrace{\left(\epsilon_\infty - \frac{\omega_p^2}{\omega^2 + i\gamma_D\omega}\right)}_{=:\epsilon(\omega)} \breve{\vec{E}} = 0. \tag{3.6}$$

Here, ϵ_∞ describes the non-dispersive background, $\omega_p = \sqrt{\frac{ne^2}{m_e}}$ is the plasma frequency and γ_D denotes the damping. The effective dielectric function is given by

$$\boxed{\text{Drude Model} \qquad \breve{\epsilon}(\omega) = \epsilon_\infty - \frac{\omega_p^2}{\omega\left(\omega + i\gamma_D\right)}} \tag{3.7}$$

and usually called Drude model. Similarly to the Lorentz model, the Drude model does not serve well as an ab-initio description of materials. Instead, it can be used
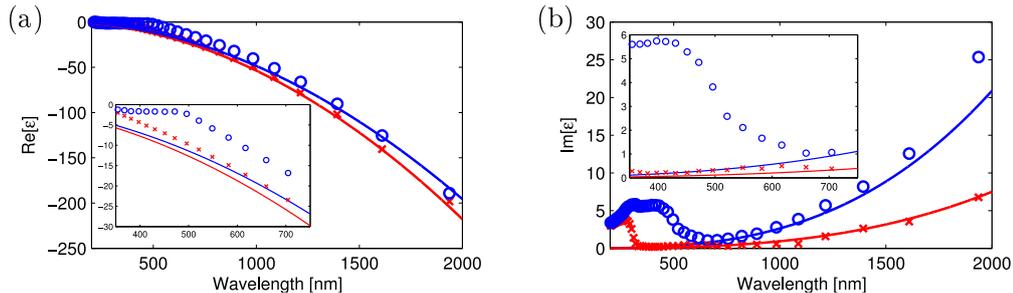
Figure 3.2: *Experimental data for the dielectric functions of gold and silver from Ref. [25]. The red crosses correspond to the data for silver while the blue circles denote measurements for gold. (a) shows the real part of ϵ, while (b) displays the imaginary part. The solid lines are Drude fits with parameters given in Tab. 3.1. The insets show a magnification around the visible spectrum.*

	Silver (Ag)	**Gold (Au)**
Plasma Frequency $\omega_{\mathbf{p}}$	$1.39 \times 10^{16} \mathrm{s}^{-1}$	$1.38 \times 10^{16} \mathrm{s}^{-1}$
Damping $\gamma_{\mathbf{D}}$	$3.23 \times 10^{13} \mathrm{s}^{-1}$	$1.08 \times 10^{14} \mathrm{s}^{-1}$

Table 3.1: *Values used to model the dispersion relation of gold and silver via a Drude model.*

to approximate the dielectric function of metals in a certain frequency range. The shape of Eq.(3.7) is plotted in Fig. 3.1(b). In order to check how well the Drude model describes real metals, we compare with data from experimental measurements for gold and silver films [25]. The symbols in Fig. 3.2 denote the experimental data, while the solid lines are fits corresponding to the values in Tab. 3.1. The fits presented in Fig. 3.2 were conducted to achieve an optimal description over the whole spectral range. It becomes obvious that both silver and gold are reasonably well described in the infrared region. In the visible, however, especially the dispersion of gold differs significantly due to intraband transitions. In order to accurately model those metals in the visible spectrum there are multiple options. One can either fit the Drude model only to a smaller spectral range which will usually improve the accuracy significantly. Alternatively, one can employ a combination of the Drude and the Lorentz model. Recently, even more sophisticated models were proposed in order to describe the visible and ultraviolet spectrum of metals more accurately [26].

3.1.3 Derivation of the ADEs

Since we aim to solve Maxwell's equations in the time-domain, a Fourier transform of $\epsilon(\omega)$ is required. Unfortunately, a straight-forward transform only recovers the convolution integrals, which render Maxwell's equations intractable. Instead, we have

to rearrange the equations and introduce auxiliary fields in order to obtain a purely differential equation in the time-domain. We start from Eq. (2.3d) without the current density,

$$-i\omega\breve{\epsilon}\vec{\breve{E}} = \nabla \times \vec{\breve{H}}, \tag{3.8}$$

in which we insert the Lorentz model (3.2) as well as the Drude model (3.7). Some algebraic transformations then yield

$$-i\omega\vec{\breve{E}} = \frac{1}{\epsilon_\infty}\left[\nabla \times \vec{\breve{H}} - \underbrace{\frac{i\omega\omega_D^2}{\omega\left(\omega - i\gamma_D\right)}\vec{\breve{E}}}_{=:\vec{\breve{j}}_D} - \underbrace{\frac{-i\omega\Delta\epsilon\,\omega_L^2}{\omega_L^2 - i\gamma_L - \omega^2}\vec{\breve{E}}}_{=:\vec{\breve{j}}_L}\right]. \tag{3.9}$$

Here, we have introduced new variables $\vec{\breve{j}}_D$ and $\vec{\breve{j}}_L$ which are called polarization currents. Looking at the two contributions separately, we can rewrite the expression for $\vec{\breve{j}}_D$ as

$$\left(i\omega + \gamma_D\right)\vec{\breve{j}}_D = \omega_D^2\vec{\breve{E}},$$

while the equation for $\vec{\breve{j}}_L$ now takes the form

$$\left(\omega_L^2 - i\gamma_L\omega - \omega^2\right)\vec{\breve{j}}_L = -i\omega\Delta\epsilon\,\omega_L^2\vec{\breve{E}}.$$

Finally, applying a Fourier transformation to both and introducing the auxiliary field $\vec{k}_L = \frac{\partial}{\partial t}\vec{j}_L$ results in the following set of coupled equations

$$\frac{\partial}{\partial t}\vec{E} = \frac{1}{\epsilon_\infty}\left[\nabla \times \vec{H} - \vec{j}_D - \vec{j}_L\right], \tag{3.10a}$$

$$\frac{\partial}{\partial t}\vec{j}_D = -\gamma_D\vec{j}_D + \omega_D^2\vec{E}, \tag{3.10b}$$

$$\frac{\partial}{\partial t}\vec{j}_L = \vec{k}_L, \tag{3.10c}$$

$$\frac{\partial}{\partial t}\vec{k}_L = -\gamma_L\vec{k}_L - \omega_L^2\vec{j}_L + \omega_L^2\Delta\epsilon\frac{\partial}{\partial t}\vec{E}. \tag{3.10d}$$

As can be seen, we need one ADE per field component to describe a material according to the Drude model, while the Lorentz model requires two ADEs. Most importantly, in both cases, there no longer is a convolution integral to solve.

3.2 Perfectly Matched Layers

In principle, Maxwell's equations as stated in Eq. (2.1) have an infinite number of solutions. To narrow them down to a unique solution, one needs to specify two more things. The first ingredient are initial conditions, which impress the fields at an initial time. The second part are boundary conditions, which define the fields on a surface enclosing a given domain. The latter is particularly important when trying to solve
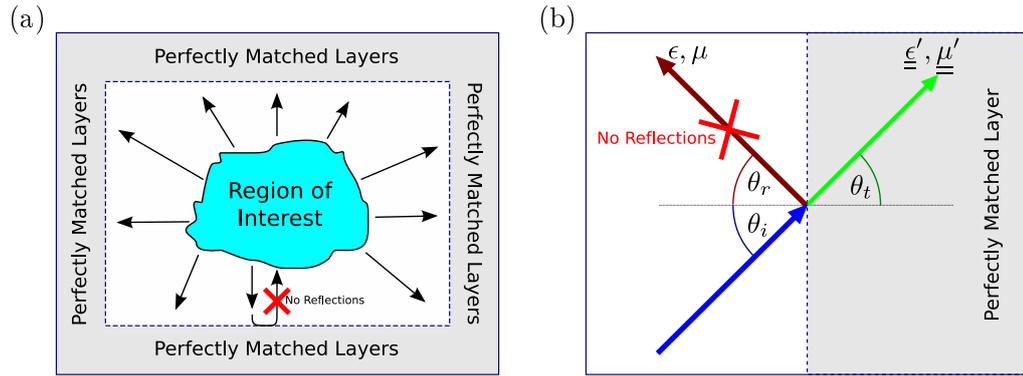
Figure 3.3: *Schematic sketch of the perfectly matched layers. (a) shows a system, where the region of interest is surrounded by perfectly matched layers. (b) shows a closeup of an incident plane wave on an individual layer. The parameters $\underline{\underline{\epsilon}}'$ and $\underline{\underline{\mu}}'$ must be chosen such that the reflection coefficient is zero for all angles of incidence.*

Maxwell's equations by means of numerical methods. Then, the computational domain has to be finite and one needs to make sure that the applied boundary conditions correspond to the physical system to solve. Of all the possible boundary conditions, one of the most practically relevant are so-called absorbing boundary conditions (ABCs). They insure that any radiation hitting the boundary is absorbed and never returns to the computational domain. Therefore, one can calculate very large open systems with dramatically reduced computational effort. Generally, we differentiate between true ABCs, which are applied at the boundaries, and absorbing materials which surround the computational region. A lot of research has gone into both methods, but in the recent years, the second method has established itself as the more practical one. As sketched in Fig. 3.3(a), one needs to implement an absorbing material around the physical system, which shows zero reflection, regardless of the frequency or angle of incidence. Such a material is called a perfectly matched layer (PML) and there is a variety of methods to implement it [27]. Here, we will concentrate on two different implementations.

3.2.1 Uniaxial Perfectly Matched Layers

As a naïve idea, one might try to simply surround the computational domain with an isotropic absorbing medium. To evaluate this, we consider a simple half-space geometry as sketched in Fig. 3.3(b). On the left side, we assume an isotropic and lossless medium with a purely real refractive index n_L. The right side is supposed to absorb radiation and is therefore assumed to have a complex ϵ' as well as a complex μ'. Considering a plane wave impinging under the angle θ_i, we know the reflection coefficients given by Fresnel's formulae (2.19) and (2.20). Due to the explicit dependence on the angle θ_i,

it is not possible to find an absorbing ϵ' and μ' such that $r_s = r_p = 0$ for all angles θ_i. In the special case of normal incidence ($\theta_i = 0$), the formulae for the two polarizations are identical and read

$$r = \frac{\sqrt{\mu/\epsilon} - \sqrt{\mu'/\epsilon'}}{\sqrt{\mu/\epsilon} + \sqrt{\mu'/\epsilon'}}.$$

In order to get the reflectivity to vanish we need to fulfill

$$\sqrt{\frac{\mu}{\epsilon}} = \sqrt{\frac{\mu'}{\epsilon'}} = \sqrt{\frac{|\mu'|\, e^{i\varphi_{\mu'}}}{|\epsilon'|\, e^{i\varphi_{\epsilon'}}}},$$

which can be realized by choosing $|\mu'| = \mu$, $|\epsilon'| = \epsilon$ and $\varphi_{\epsilon'} = \varphi_{\mu'}$. In this case, we speak of impedance matching.

In order to achieve such matching for any angle θ_i, we have to reconsider the assumption of isotropy and continue with an anisotropic medium on the right side. As we will demonstrate, a diagonal tensor is sufficient and—due to the rotational symmetry around the x-axis—only a uniaxial form is required. Therefore, inside the PML region ($x > 0$), we define

$$\underline{\underline{\epsilon}}' = \epsilon' \underline{\underline{\Lambda}} \quad \text{and} \quad \underline{\underline{\mu}}' = \mu' \underline{\underline{\Lambda}}$$

with

$$\underline{\underline{\Lambda}} = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & b \end{pmatrix}.$$

Despite the anisotropy, the plane wave solutions for s-polarization in the PML-region still take the from

$$\vec{E}_T = T_s \hat{e}_z e^{i(\vec{k}_T \vec{r} - \omega t)}$$

with the difference that the magnetic field now reads [27]

$$\vec{H} = \frac{T_s}{\omega \mu'} \left(\frac{k_{Ty}}{a} \hat{e}_x - \frac{k_{Tx}}{b} \hat{e}_y \right) e^{i(\vec{k}_T \vec{r} - \omega t)}$$

and the dispersion relation is changed to

$$ab^2 \epsilon' \mu' \omega^2 = a k_{Tx}^2 + b k_{Ty}^2. \tag{3.11}$$

Employing the boundary conditions (2.12) at $x = 0$ then yields

$$\frac{k_{Ix}}{\omega \mu} \left(1 - R_s \right) = \frac{k_{Tx}}{\omega \mu' b}$$

from which we obtain

$$R_s = 1 - \frac{\mu}{\mu'} \frac{1}{b} \frac{k_{Tx}}{k_{Ix}}. \tag{3.12}$$

From the dispersion relation (3.11) and using $k_{Iy} = k_{Ty}$ we derive

$$k_{Tx}^2 = b^2 \epsilon' \mu' \omega^2 - \frac{b}{a} k_{Iy}^2 = b^2 \left(\frac{\epsilon' \mu'}{\epsilon \mu} k_I^2 - \frac{1}{ab} k_{Iy}^2 \right).$$

If we choose

$$\epsilon' = \epsilon, \quad \mu' = \mu \quad \text{and} \quad a^{-1} = b, \tag{3.13}$$

then $k_{Tx}^2 = b^2 k_{Ix}^2$ leads to a vanishing reflectivity (3.12), independently of the choice of b. The transmitted wave is then given by

$$\vec{E}_T = T_s \hat{e}_z e^{i(\vec{k}_T \vec{r} - \omega t)} = T_s \hat{e}_z e^{i(b k_{Ix} x + k_{Iy} y - \omega t)}.$$

In order to fulfill the requirement of absorption, a suitable choice for b is

$$b = a^{-1} = s_x = 1 - \frac{\sigma_x}{i\omega},$$

where $\sigma_x > 0$ is a real constant determining the loss. With this choice, we observe that the transmitted plane wave in the PML region evolves as

$$\vec{E}_T = T_s \hat{e}_z e^{i \vec{k}_i \vec{r} - \omega t} e^{-\sigma_x n \cos(\theta_i) x}$$

and is therefore exponentially damped as it propagates in x-direction. We also realize that it is necessary to make s_x dispersive so that the damping becomes frequency independent.

As in the isotropic case, it should be noted that the result for p-polarization can be obtained by swapping ϵ' and μ'. However, since the matching conditions Eq. (3.13) are invariant under this transformation, our result is valid for both polarizations.

Generalization to Three Dimension

After finding the uniaxial tensor in x-direction, a generalization to three dimensions is straight-forwardly defined as

$$\underline{\underline{\Lambda}} = \begin{pmatrix} s_x^{-1} & 0 & 0 \\ 0 & s_x & 0 \\ 0 & 0 & s_x \end{pmatrix} \begin{pmatrix} s_y & 0 & 0 \\ 0 & s_y^{-1} & 0 \\ 0 & 0 & s_y \end{pmatrix} \begin{pmatrix} s_z & 0 & 0 \\ 0 & s_z & 0 \\ 0 & 0 & s_z^{-1} \end{pmatrix} = \begin{pmatrix} \frac{s_y s_z}{s_x} & 0 & 0 \\ 0 & \frac{s_x s_z}{s_y} & 0 \\ 0 & 0 & \frac{s_x s_y}{s_z} \end{pmatrix}.$$

The tensor elements are now given by

$$s_k = 1 - \frac{\sigma_k}{i\omega}, \quad k = x, y, z. \tag{3.14}$$

This tensor is technically no longer uniaxial, but rather biaxial. Still, this implementation goes under the name uniaxial perfectly matched layers (UPML).

Derivation of the ADEs

As mentioned, the tensor elements (3.14) are frequency dependent. This puts us in a similar position as with the dispersive dielectric function discussed in section 3.1.3. Inserting the tensor into Maxwell's equations in the frequency domain (2.3) yields

$$i\omega\epsilon\underline{\underline{\Lambda}}\breve{\vec{E}} = \nabla \times \breve{\vec{H}}, \tag{3.15a}$$

$$-i\omega\mu\underline{\underline{\Lambda}}\breve{\vec{H}} = \nabla \times \breve{\vec{E}}. \tag{3.15b}$$

We start by only looking at the first component of Eq. (3.15) and rewriting it slightly to

$$i\omega\epsilon\breve{E}_x = -\frac{\partial}{\partial y}\breve{H}_z + \frac{\partial}{\partial z}\breve{H}_y - \underbrace{i\omega\epsilon\left(\frac{s_y s_z}{s_x} - 1\right)\breve{E}_x}_{\breve{J}_x}.$$

Concentrating on the expression in brackets, we can insert Eq. (3.14) and reduce it to

$$\frac{s_y s_z}{s_x} - 1 = \frac{1}{i\omega + \sigma_x}\left(\sigma_y + \sigma_z - \sigma_x + \frac{\sigma_y \sigma_z}{i\omega}\right).$$

With this, the polarization current $\breve{J}_x$ becomes

$$\breve{J}_x = \frac{i\omega\epsilon}{i\omega + \sigma_x}\left(\sigma_y + \sigma_z - \sigma_x + \frac{\sigma_y \sigma_z}{i\omega}\right)\breve{E}_x.$$

By introducing a new variable $\breve{P}_x = \breve{J}_x - \epsilon(\sigma_y + \sigma_z - \sigma_x)\breve{E}_x$ we obtain the expression

$$i\omega\breve{P}_x = -\sigma_x\breve{P}_x + \epsilon\left(\sigma_x^2 + \sigma_y\sigma_z - \sigma_y\sigma_x - \sigma_z\sigma_x\right)\breve{E}_x$$

which can now be easily transformed back to the time-domain. The two coupled equations then read

$$\frac{\partial}{\partial t}E_x = -\frac{\partial}{\partial y}H_z + \frac{\partial}{\partial z}H_y - \epsilon\left(\sigma_y + \sigma_z - \sigma_x\right)E_x - P_x,$$

$$\frac{\partial}{\partial t}P_x = -\sigma_x P_x + \epsilon\left(\sigma_x^2 + \sigma_y\sigma_z - \sigma_y\sigma_x - \sigma_z\sigma_x\right)E_x.$$

All other components can be obtained by simple permutation of the indices. Thus, we have derived a set of auxiliary differential equations which allows us to include the dispersive behavior of the PMLs in the time domain.

3.2.2 Stretched Coordinates CFS-PMLs

The UPML formulation discussed so far is a simple, yet very effective method to implement absorbing boundary conditions. There are, however, two minor problems:

- The termination of dispersive or anisotropic material requires an adjustment of the PML formulation,

- The absorption of evanescent waves leaking into the PMLs is not optimal [27, 28].

To overcome the first problem, we discuss a different implementation of PMLs, called stretched coordinates formulation. In this case, we do not modify the material parameters but rather change Maxwell's equations inside the absorbing layers. Starting from Eq. (2.3) without free charges and currents, we modify the spatial differentiation operators to

$$\nabla_s \cdot \breve{\vec{D}}(\vec{r}, \omega) = 0,$$

$$\nabla_s \cdot \breve{\vec{B}}(\vec{r}, \omega) = 0,$$

$$\nabla_s \times \breve{\vec{E}}(\vec{r}, \omega) = i\omega \breve{\vec{B}}(\vec{r}, \omega),$$

$$\nabla_s \times \breve{\vec{H}}(\vec{r}, \omega) = -i\omega \breve{\vec{D}}(\vec{r}, \omega),$$

where

$$\nabla_s = \left(\frac{1}{s_x} \frac{\partial}{\partial x}, \frac{1}{s_y} \frac{\partial}{\partial y}, \frac{1}{s_z} \frac{\partial}{\partial z} \right)^T.$$

Similarly to the previous section, one can prove that the stretching factors do not lead to the reflection of a plane wave [27, 28]. This is even true for dispersive or anisotropic media, since the formulation is independent of the material properties. It should be noted, that the UPML formulation and the stretched-coordinate formulation are mathematically identical for isotropic media [29].

In order to also mitigate the problems with evanescent waves, we introduce a particular choice of stretching factors s_i, namely

$$s_k = \kappa_k - \frac{\sigma_k}{\alpha_k + i\omega}, \quad k = x, y, z.$$

Due to the newly introduced factor α_k, this formulation is called complex frequency-shifted PMLs (CFS-PMLs) [28]. In contrast to Eq. (3.14), we have introduced a second new coefficient κ_k, which corresponds to a real stretching of the coordinates. As is shown in Ref. [28], this factor can help to improve the numerical performance of PMLs dramatically. The only disadvantage when compared to the UPMLs is that an implementation via a single ADE per field component is no longer possible. Instead, we need a set of two ADEs, which increases computational effort and memory consumption. We will skip the derivation, which is very analogous to the one in Sec. 3.2.1. Instead, we only state the final expression for E_x and the two auxiliary fields as

$$\frac{\partial}{\partial t} E_x = \frac{1}{\epsilon} \left(\frac{1}{\kappa_y} \frac{\partial}{\partial y} H_z - \frac{1}{\kappa_z} \frac{\partial}{\partial z} H_y - F_{xy} + F_{xz} \right), \tag{3.16}$$

$$\frac{\partial}{\partial t} F_{xy} = \frac{\sigma_y}{\kappa_y^2} \frac{\partial}{\partial y} H_z - \left(\alpha_y + \frac{\sigma_y}{\kappa_y} \right) F_{xy}, \tag{3.17}$$

$$\frac{\partial}{\partial t} F_{xz} = \frac{\sigma_z}{\kappa_z^2} \frac{\partial}{\partial z} H_y - \left(\alpha_z + \frac{\sigma_z}{\kappa_z} \right) F_{xz}. \tag{3.18}$$

All other components can again be obtained by proper permutation of the indices. We will return to the numerical implementation and careful analysis of the performance in Sec. 4.3.3 and Sec. 5.4.

4 The Finite-Difference Time-Domain Method

As alluded to in the introduction, analytical solutions to partial differential equations (PDEs) like Maxwell's equations are notoriously difficult to obtain for any non-trivial system. If there are no suitable approximations which facilitate analytical treatment, one usually has to resort to numerical methods for finding solutions to those PDEs. If a solution in the time-domain is desired, a typical procedure is to first discretize the system in space. This reduces the set of PDEs to a large system of ordinary differential equations (ODEs), which can then be solved in time by means of standard numerical integrators. This procedure is known as the method of lines (MOL).

In this chapter, we discuss the simplest spatial discretization possible, namely finite differences on a rectangular uniform mesh. It will be demonstrated, that Maxwell's equations have properties which lead to a particular spatial arrangement, known as a staggered grid. Using finite differences for the time integration as well leads us to the famous finite-difference time-domain (FDTD) method. We will thoroughly study the stability and accuracy of this method, before we discuss possible extensions and improvements. The chapter concludes with a number of basic numerical experiments in order to verify our implementation of the method.

4.1 Discretization with Finite Differences

In order to solve any PDE numerically, one needs to discretize its continuous fields to reduce the problem to a finite number of unknowns. Possibly the simplest method to do so is to employ a uniform grid as sketched in Fig. 4.1. Starting from this discretization, one can expand an arbitrary field $F(x)$ around the points $x_0 - \Delta x$ and $x_0 + \Delta x$ using Taylor's theorem as

$$F(x_0 \pm \Delta x) = F(x_0) \pm F'(x_0)\Delta x + \frac{1}{2}F''(x_0)\Delta x^2 \pm O(\Delta x^3).$$

Combining the two equations allows to solve for $F'(x_0)$ and to obtain an approximation of the first derivative of F at x_0. In principle, three different approximations are
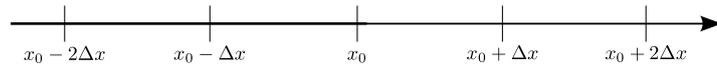
Figure 4.1: *Sketch of a uniform grid in one dimension.*

possible:

$$\text{Forward Differences:} \qquad F'(x_0) = \frac{F(x_0 + \Delta x) - F(x_0)}{\Delta x} + O(\Delta x),$$

$$\text{Central Differences:} \qquad F'(x_0) = \frac{F(x_0 + \Delta x) - F(x_0 - \Delta x)}{2\Delta x} + O(\Delta x^2),$$

$$\text{Backward Differences:} \qquad F'(x_0) = \frac{F(x_0) - F(x_0 - \Delta x)}{\Delta x} + O(\Delta x).$$

From looking at the truncation order, we observe that the central difference yields a significantly better convergence for $\Delta x \to 0$. For this reason, we will continue working exclusively with central differences.

4.1.1 The One-Dimensional Case

In order to understand some of the basic features of the discretized Maxwell equations, we start from the one-dimensional formulation (2.15) without the source terms:

$$\partial_t H_y = \frac{1}{\mu} \partial_x E_z,$$

$$\partial_t E_z = \frac{1}{\epsilon} \partial_x H_y.$$

Employing a uniform grid as sketched in Fig. 4.1 and discretizing the spatial derivatives by central differences yields the semi-discrete system

$$\partial_t H_y(x,t) = \frac{1}{\mu} \frac{E_z(x + \Delta x, t) - E_z(x - \Delta x, t)}{2\Delta x},$$

$$\partial_t E_z(x,t) = \frac{1}{\epsilon} \frac{H_y(x + \Delta x, t) - H_y(x - \Delta x, t)}{2\Delta x}.$$

By close inspection of these semi-discrete equations, one can observe that the grid decouples into two unrelated subgrids. The electrical field at point x_0 is not related to the electrical fields at points $x_0 + n\Delta x$ for odd n and vice versa. The same also holds for the magnetic field. This behavior is known as even-odd decoupling and indicates that above discretization is far from ideal. Instead, it makes sense to pick one of the two subgrids and neglect all unrelated field points. This leads us to the staggered grid as sketched in Fig. 4.2. To keep our notation consistent with the literature, we rescale the distance between two neighboring gridpoints to $\frac{\Delta x}{2}$ and introduce the
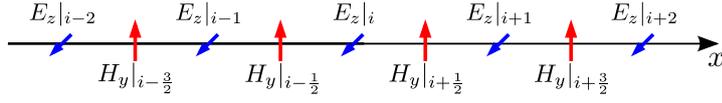
Figure 4.2: *Sketch of a staggered grid in one dimension. The blue arrows indicate a gridpoint of an E_z field component, while the red arrows denote the positions of the H_y field.*

notation $F|_i = F(i\Delta x, t)$. Here, all electrical field points have integer indices, while the magnetic fields are located at half-integer indices as shown in Fig. 4.2. The semi-discretized Maxwell's equations in this notation then read

$$\partial_t H_y|_{i+\frac{1}{2}} = \frac{1}{\mu|_{i+\frac{1}{2}}} \frac{E_z|_{i+1} - E_z|_i}{\Delta x},$$

$$\partial_t E_z|_i = \frac{1}{\epsilon|_i} \frac{H_y|_{i+\frac{1}{2}} - H_y|_{i-\frac{1}{2}}}{\Delta x},$$

where we also discretized the material parameters μ and ϵ.

So far, we have only discretized the spatial derivatives and derived a coupled system of ODEs in time. However, nothing prevents us from applying finite differences to the time derivatives as well. By replacing $\frac{\partial}{\partial t}$ with central differences, we again observe an even-odd decoupling, where the fields at a certain point in time are unrelated to the same fields at a neighboring point in time. Therefore, we again decide to drop every other timestep and end up with a scheme called leap-frogging. The fully discretized equations can be written as

$$\frac{H_y|_{i+\frac{1}{2}}^{n+\frac{1}{2}} - H_y|_{i+\frac{1}{2}}^{n-\frac{1}{2}}}{\Delta t} = \frac{1}{\mu|_{i+\frac{1}{2}}} \frac{E_z|_{i+1}^n - E_z|_i^n}{\Delta x}, \tag{4.2a}$$

$$\frac{E_z|_i^{n+1} - E_z|_i^n}{\Delta t} = \frac{1}{\epsilon|_i} \frac{H_y|_{i+\frac{1}{2}}^{n+\frac{1}{2}} - H_y|_{i-\frac{1}{2}}^{n+\frac{1}{2}}}{\Delta x}, \tag{4.2b}$$

where we extended our notation to $F|_i^n = F(i\Delta x, n\Delta t)$. In order to obtain an explicit update scheme, we can solve Eqs. (4.2) for the future timestep and obtain

---
FDTD Update Equations (1D)

$$H_y|_{i+\frac{1}{2}}^{n+\frac{1}{2}} = H_y|_{i+\frac{1}{2}}^{n-\frac{1}{2}} + \frac{1}{\mu|_{i+\frac{1}{2}}} \frac{\Delta t}{\Delta x} \left(E_z|_{i+1}^n - E_z|_i^n \right), \tag{4.3a}$$

$$E_z|_i^{n+1} = E_z|_i^n + \frac{1}{\epsilon|_i} \frac{\Delta t}{\Delta x} \left(H_y|_{i+\frac{1}{2}}^{n+\frac{1}{2}} - H_y|_{i-\frac{1}{2}}^{n+\frac{1}{2}} \right). \tag{4.3b}$$

From this result, one can directly observe how a new value of the electric field $E_z|_i^{n+1}$ is generated from the old value $E_z|_i^n$ at the same position and the two neighboring
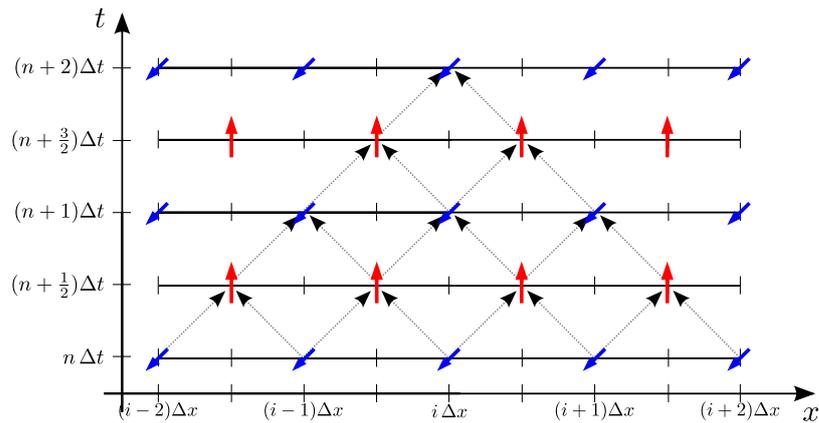
Figure 4.3: *Sketch of the leap-frog integration in time as used by the FDTD method. The blue arrows indicate a gridpoint of an E_z field component, while the red arrows denote the positions of the H_y field.*

magnetic field points $H_y\big|_{n+\frac{1}{2}}^{i+\frac{1}{2}}$ and $H_y\big|_{n+\frac{1}{2}}^{i-\frac{1}{2}}$ at an in-between time. The procedure is sketched in Fig. 4.3 and commonly referred to as the one-dimensional finite-difference time-domain (FDTD) method.

4.1.2 Higher Dimensions and the Yee Cube

Similarly to the one-dimensional case, explicit update schemes can also be derived in two and three dimensions. First, we generalize our notation to

$$F\big|_{i,j,k}^n = F(i\Delta x, j\Delta y, k\Delta z, n\Delta t).$$

Next, we employ central differences to discretize all derivatives and again observe odd-even decoupling which leads us to staggered grids.

Two-Dimensional Systems

In two dimensions, we can deal with the two different polarizations separately. Since the two sets of equations (2.13), (2.14) are related by a simple transformation, the same holds for the grid, as shown in Fig. 4.4. Inserting central differences for the spatial and temporal derivatives will then result in the explicit update equations for
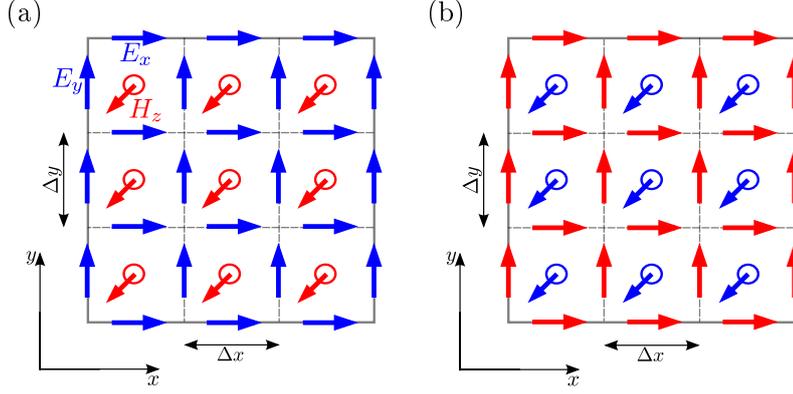
Figure 4.4: *Schematic layout of the grid positions for a FDTD scheme in two dimensions. The positions of the fields in (a) TE and (b) TM polarization are related by swapping the electric and magnetic field components. Blue arrows indicate the electric field, while red arrows denote magnetic field components.*

TE polarization as

$$
\boxed{
\begin{aligned}
&\textbf{FDTD Update Equations (2D, TE Polarization)} \\[4pt]
E_x\big|_{i,j+\frac{1}{2}}^{n+1} &= E_x\big|_{i+\frac{1}{2},j}^{n+1} + \frac{1}{\epsilon\big|_{i,j+\frac{1}{2}}}\frac{\Delta t}{\Delta y}\left(H_z\big|_{i,j+1}^{n+\frac{1}{2}} - H_z\big|_{i,j}^{n+\frac{1}{2}}\right), &&(4.4a)\\[6pt]
E_y\big|_{i+\frac{1}{2},j}^{n+1} &= E_y\big|_{i,j+\frac{1}{2}}^{n+1} - \frac{1}{\epsilon\big|_{i+\frac{1}{2},j}}\frac{\Delta t}{\Delta x}\left(H_z\big|_{i+1,j}^{n+\frac{1}{2}} - H_z\big|_{i,j}^{n+\frac{1}{2}}\right), &&(4.4b)\\[6pt]
H_z\big|_{i,j}^{n+\frac{1}{2}} &= H_z\big|_{i,j}^{n-\frac{1}{2}} + \frac{\Delta t}{\mu\big|_{i,j}}\left(\frac{E_x\big|_{i,j+\frac{1}{2}}^{n} - E_x\big|_{i,j-\frac{1}{2}}^{n}}{\Delta y} - \frac{E_y\big|_{i+\frac{1}{2},j}^{n} - E_y\big|_{i-\frac{1}{2},j}^{n}}{\Delta z}\right). &&(4.4c)
\end{aligned}
}
$$

Equivalently, the scheme for TM polarization reads

$$
\boxed{
\begin{aligned}
&\textbf{FDTD Update Equations (2D, TM Polarization)} \\[4pt]
H_x\big|_{i,j+\frac{1}{2}}^{n+1} &= H_x\big|_{i+\frac{1}{2},j}^{n+1} - \frac{1}{\mu\big|_{i,j+\frac{1}{2}}}\frac{\Delta t}{\Delta y}\left(E_z\big|_{i,j+1}^{n+\frac{1}{2}} - E_z\big|_{i,j}^{n+\frac{1}{2}}\right), &&(4.5a)\\[6pt]
H_y\big|_{i+\frac{1}{2},j}^{n+1} &= H_y\big|_{i,j+\frac{1}{2}}^{n+1} + \frac{1}{\mu\big|_{i+\frac{1}{2},j}}\frac{\Delta t}{\Delta x}\left(E_z\big|_{i+1,j}^{n+\frac{1}{2}} - E_z\big|_{i,j}^{n+\frac{1}{2}}\right), &&(4.5b)\\[6pt]
E_z\big|_{i,j}^{n+\frac{1}{2}} &= E_z\big|_{i,j}^{n-\frac{1}{2}} + \frac{\Delta t}{\epsilon\big|_{i,j}}\left(\frac{H_x\big|_{i,j+\frac{1}{2}}^{n} - H_x\big|_{i,j-\frac{1}{2}}^{n}}{\Delta y} - \frac{H_y\big|_{i+\frac{1}{2},j}^{n} - H_y\big|_{i-\frac{1}{2},j}^{n}}{\Delta z}\right). &&(4.5c)
\end{aligned}
}
$$

It is important to note, that the material parameters ϵ and μ are also staggered, which complicates the treatment of anisotropic or nonlinear media.
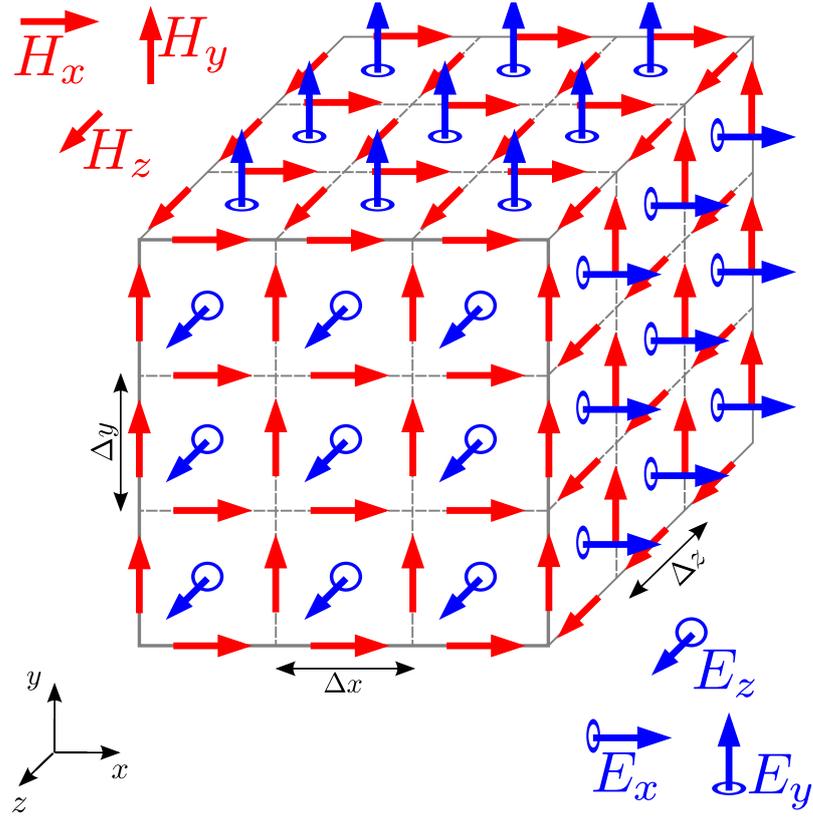
Figure 4.5: *Yee grid, the three-dimensional spatial arrangement of the field components in the FDTD method.*

Three-Dimensional Systems

For the fully three-dimensional case, all six field components are staggered with respect to each other. Part of the grid is sketched in Fig. 4.5 and this type of arrangement is often called Yee cube after Kane Yee [30]. The discretization procedure is performed analogous to the one- and two-dimensional case and leads to the semi-discrete system

$$
\partial_t E_x\big|_{i,j+\frac{1}{2},k+\frac{1}{2}} = \frac{1}{\epsilon\big|_{i,j+\frac{1}{2},k+\frac{1}{2}}} \left(\frac{H_z\big|_{i,j+1,k+\frac{1}{2}} - H_z\big|_{i,j,k+\frac{1}{2}}}{\Delta y} - \frac{H_y\big|_{i,j+\frac{1}{2},k+1} - H_y\big|_{i,j+\frac{1}{2},k}}{\Delta z} \right),
$$

$$
\partial_t E_y\big|_{i+\frac{1}{2},j,k+\frac{1}{2}} = \frac{1}{\epsilon\big|_{i+\frac{1}{2},j,k+\frac{1}{2}}} \left(\frac{H_x\big|_{i+\frac{1}{2},j,k+1} - H_x\big|_{i+\frac{1}{2},j,k}}{\Delta z} - \frac{H_z\big|_{i+1,j,k+\frac{1}{2}} - H_z\big|_{i,j,k+\frac{1}{2}}}{\Delta x} \right),
$$

$$
\partial_t E_z\big|_{i+\frac{1}{2},j+\frac{1}{2},k} = \frac{1}{\epsilon\big|_{i+\frac{1}{2},j+\frac{1}{2},k}} \left(\frac{H_y\big|_{i+1,j+\frac{1}{2},k} - H_y\big|_{i,j+\frac{1}{2},k}}{\Delta x} - \frac{H_x\big|_{i+\frac{1}{2},j+1,k} - H_x\big|_{i+\frac{1}{2},j,k}}{\Delta y} \right),
$$

$$\partial_t H_x|_{i+\frac{1}{2},j,k} = \frac{1}{\mu|_{i+\frac{1}{2},j,k}} \left(\frac{E_y^{i+\frac{1}{2},j,k+\frac{1}{2}} - E_y^{i+\frac{1}{2},j,k-\frac{1}{2}}}{\Delta z} - \frac{E_z^{i+\frac{1}{2},j+\frac{1}{2},k} - E_z^{i+\frac{1}{2},j-\frac{1}{2},k}}{\Delta y} \right),$$

$$\partial_t H_y|_{i,j+\frac{1}{2},k} = \frac{1}{\mu|_{i,j+\frac{1}{2},k}} \left(\frac{E_z^{i+\frac{1}{2},j+\frac{1}{2},k} - E_z^{i-\frac{1}{2},j+\frac{1}{2},k}}{\Delta x} - \frac{E_x^{i,j+\frac{1}{2},k+\frac{1}{2}} - E_x^{i,j+\frac{1}{2},k-\frac{1}{2}}}{\Delta z} \right),$$

$$\partial_t H_z|_{i,j,k+\frac{1}{2}} = \frac{1}{\mu|_{i,j,k+\frac{1}{2}}} \left(\frac{E_x^{i,j+\frac{1}{2},k+\frac{1}{2}} - E_x^{i,j-\frac{1}{2},k+\frac{1}{2}}}{\Delta y} - \frac{E_y^{i+\frac{1}{2},j,k+\frac{1}{2}} - E_y^{i-\frac{1}{2},j,k+\frac{1}{2}}}{\Delta x} \right).$$

The discretization in time can be performed as in the one-dimensional case to obtain the fully explicit update equations. For brevity, we only state the explicit update equation for the E_x as

$$E_x|_{i,j+\frac{1}{2},k+\frac{1}{2}}^{n+1} = E_x|_{i,j+\frac{1}{2},k+\frac{1}{2}}^{n}$$

$$+ \frac{\Delta t}{\epsilon|_{i,j+\frac{1}{2},k+\frac{1}{2}}} \left(\frac{H_z|_{i,j+1,k+\frac{1}{2}}^{n+\frac{1}{2}} - H_z|_{i,j,k+\frac{1}{2}}^{n+\frac{1}{2}}}{\Delta y} - \frac{H_y|_{i,j+\frac{1}{2},k+1}^{n+\frac{1}{2}} - H_y|_{i,j+\frac{1}{2},k}^{n+\frac{1}{2}}}{\Delta z} \right). \quad (4.6)$$

From the spatial arrangement displayed in Fig. 4.5, we observe how the structure of Maxwell's equations is embedded in the grid. As an example, let us consider an arbitrary E_z-component. The change of this field value in time is given by the magnetic fields H_x and H_y in a loop around E_z. This behavior corresponds to Maxwell's equations in their integral form, where the change of the electric flux through a loop is given by the line integral over the magnetic field around the loop. Similar considerations apply to all other field components.

Finally, it should be noted that we only discretized the curl equations so far. As argued in the Sec. 2.1, the divergence conditions remain fulfilled if they are valid initially. This remains true for the FDTD scheme as proven explicitly in Ref. [28].

4.2 Convergence, Stability and Accuracy

A major point and a condition sine qua non for every numerical method is the proof of convergence. Unfortunately, such proofs are often riddled with mathematical subtleties and rarely done in a few lines. In the interest of briefness, we will therefore skip the full mathematical treatment and rather concentrate on some of the key points. An exhaustive discussion of the convergence of finite-difference schemes can be found in Ref. [31].

Here, we will concentrate on the practical implications for Maxwell's equations. As shown in Sec. 2.4, for a homogeneous system we can find general solutions as plane waves

$$\vec{E} = \vec{E}_0 e^{i(\vec{k}\vec{r} - \omega t)} \quad \text{and} \quad \vec{H} = \vec{H}_0 e^{i(\vec{k}\vec{r} - \omega t)} \quad (4.7)$$

with the dispersion relation

$$\omega = \frac{|\vec{k}|}{n} = \frac{k}{n}.$$

Remembering the dimensionless units, in which the vacuum speed of light is $c = 1$, we can express the phase velocity of the plane waves in free space as

$$v_p = \frac{\omega}{k} = 1.$$

In order to compare these analytical solutions with our numerical ones, we insert the ansatz (4.7) into the three-dimensional update equations (4.6). To avoid confusion with the exact solution, we indicate the numerical frequency and the wavevector by $\tilde{\omega}$ and $\vec{\tilde{k}}$, respectively. This leads to the condition

$$\left[\frac{\sin\left(\tilde{\omega}\Delta t/2\right)}{c\Delta t}\right]^2 = \left[\frac{1}{\Delta x}\sin\left(\frac{\tilde{k}_x \Delta x}{2}\right)\right]^2 + \left[\frac{1}{\Delta y}\sin\left(\frac{\tilde{k}_y \Delta y}{2}\right)\right]^2 + \left[\frac{1}{\Delta z}\sin\left(\frac{\tilde{k}_z \Delta z}{2}\right)\right]^2.$$

Solving for $\tilde{\omega}$ then gives the numerical dispersion relation

$$\tilde{\omega} = \frac{2}{\Delta t}\arcsin\left(\xi\right), \tag{4.8}$$

where

$$\xi = \Delta t \sqrt{\left[\frac{1}{\Delta x}\sin\left(\frac{\tilde{k}_x \Delta x}{2}\right)\right]^2 + \left[\frac{1}{\Delta y}\sin\left(\frac{\tilde{k}_y \Delta y}{2}\right)\right]^2 + \left[\frac{1}{\Delta z}\sin\left(\frac{\tilde{k}_z \Delta z}{2}\right)\right]^2}.$$

4.2.1 Stability

As can be seen from the ansatz (4.7), the numerical frequency $\tilde{\omega}$ needs to stay purely real. Otherwise, we either find an unphysical damping or an exponential growth of the propagating waves. In order to fulfill this condition, we require

$$\xi \overset{!}{\leq} 1,$$

which immediately results in a restriction on the timestep Δt as

$$\boxed{\begin{array}{c} \text{CFL Criterion} \\ \Delta t \leq \frac{1}{\sqrt{\frac{1}{\Delta x^2} + \frac{1}{\Delta y^2} + \frac{1}{\Delta z^2}}}. \end{array}} \tag{4.9}$$

This criterion is also known as the Courant-Friedrichs-Levy (CFL) criterion. A more detailed analysis [28] proves, that any violation of this condition will indeed lead to exponentially growing waves and therefore results in an instability. Mathematically speaking, the CFL-criterion (4.9) is a necessary and sufficient condition for stability of the FDTD method [31].

4.2.2 Accuracy

Besides the stability, we are also interested in comparing the numerical dispersion relation (4.8) to the analytical case

$$\omega^2 = k_x^2 + k_y^2 + k_z^2,$$

in order to get a feeling for the errors introduced by the discretization. To proceed, we restrict our analysis to a cubic mesh, where $\Delta = \Delta x = \Delta y = \Delta z$. Further, we assume a timestep $\Delta t = s\Delta$ where s is called the Courant number and $s \leq \frac{1}{\sqrt{3}}$ is required to fulfill the CFL-criterion. Inserting these simplifications into Eq. (4.8) results in

$$\tilde{\omega} = \frac{2}{s\Delta} \arcsin \left(s \sqrt{\sin^2 \left(\frac{\tilde{k}_x \Delta}{2} \right) + \sin^2 \left(\frac{\tilde{k}_y \Delta}{2} \right) + \sin^2 \left(\frac{\tilde{k}_z \Delta}{2} \right)} \right).$$

A Taylor expansion for small Δ then gives

$$\tilde{\omega} = \sqrt{\tilde{k}_x^2 + \tilde{k}_y^2 + \tilde{k}_z^2} + O(\Delta^2)$$

and indicates that the FDTD method is accurate to second order. This was to be expected since we used central differences which are of second order. To gain a better understanding for the effect of this error, we now assume propagation along one of the grid axes, such that $\tilde{k}_x = \tilde{k}$ and $\tilde{k}_y = \tilde{k}_z = 0$. With that, Eq. (4.8) reduces to

$$\tilde{\omega} = \frac{2}{s\Delta} \arcsin \left(s \sin \left(\frac{\tilde{k}_x \Delta}{2} \right) \right).$$

Interestingly, we would regain the exact expression for $s = 1$, but due to the CFL-criterion such a "magic" timestep is prohibited. Next, we express $\tilde{k}$ as a function of the wavelength λ

$$\tilde{k} = \frac{2\pi}{\lambda} = \frac{2\pi}{N_\lambda \Delta},$$

where we have further introduced $N_\lambda = \frac{\lambda}{\Delta}$ as a measure of the number of cells per wavelength. With this, we can calculate the numerical phase velocity as

$$\tilde{v} = \frac{\tilde{\omega}}{\tilde{k}} = \frac{N_\lambda}{\pi s} \arcsin \left(s \sin \left(\frac{\pi}{N_\lambda} \right) \right).$$

In Fig. 4.6(a), we plot the numerical phase velocity error $1 - \tilde{v}$ as a function of gridpoints per wavelength N_λ. One observes that approximately 10 gridpoints per wavelength are required to reduce the phase velocity error below 1%. Around 35 points are needed to push the error below 0.1%. These numbers can serve as a rough estimate of how one should choose the discretization of a given system. However, one must not make the mistake to conclude from an error in the phase velocity to an error in the fields.
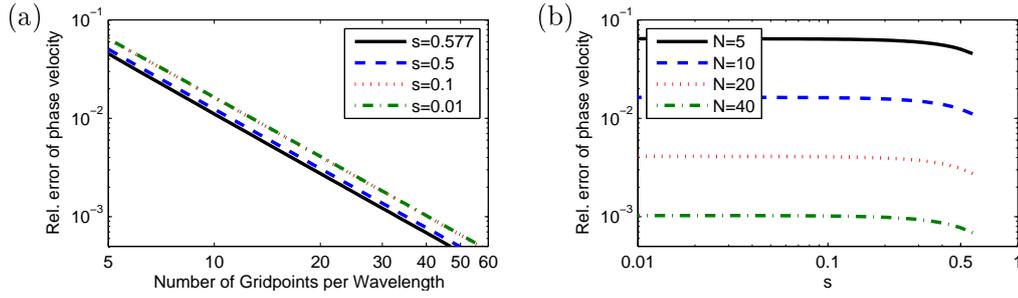
Figure 4.6: *The error in the phase velocity $\tilde{v}$ as a function of (a) the number of gridpoints per wavelength and (b) the timestep factor s. Both are double logarithmic plots and the lines in panel (a) show a slope of -2. The plots in (b) end at $s \approx 0.577$ due to the CFL criterion.*

As a simple example, we consider a plane wave with wavelength λ and phase velocity $\tilde{v} = 0.99v$. After a distance of $x = 50\lambda$ we have picked up a phase of π when compared to the exact solution. This corresponds to a relative field error of 200%.

From the data in Fig. 4.6(b), we further observe that one should choose the Courant number s as large as possible to reduce the error. This is also beneficial from a second point of view, since it increases the timestep Δt and therefore speeds up the calculation. As discussed previously, we must stay below the CFL-criterion which dictates the optimal value of s as $\frac{1}{\sqrt{3}} \approx 0.577$.

In a next step, we look at the possible anisotropy of the phase velocity by considering a $\vec{k}$-vector in the xy-plane. We therefore introduce $\tilde{k}_x = \tilde{k}\cos(\varphi)$ and $\tilde{k}_y = \tilde{k}\sin(\varphi)$, while $\tilde{k}_z = 0$. Insertion into Eq. (4.8) allows us to determine the numerical phase velocity

$$\tilde{v} = \frac{N_\lambda}{\pi s} \arcsin\left(s\sqrt{\sin^2\left(\frac{\pi}{N_\lambda}\cos(\varphi)\right) + \sin^2\left(\frac{\pi}{N_\lambda}\sin(\varphi)\right)} \right)$$

as a function of the propagation angle φ. We plot the error for different discretizations in Fig. 4.7 and observe, how the phase error is significantly reduced for propagation under 45°. In fact, if we consider propagation along the space diagonal, i.e., $\tilde{k}_x = \tilde{k}_y = \tilde{k}_z = \frac{1}{\sqrt{3}}\tilde{k}$, the phase velocity error vanishes entirely for $s = \frac{1}{\sqrt{3}}$. Unfortunately, this is of little relevance since in any realistic system, there will be propagation in multiple directions.

4.2.3 Improvements by Using Higher-Order Discretizations

As we have seen in the previous section, the propagation error is mostly determined by the number of gridpoints per wavelength. Even the lower limit of ten points per smallest wavelength in the system can be very restricting for larger three-dimensional systems. For this reason, a massive amount of research has gone into the question of how one can reduce the required number of points per wavelength. If we want to
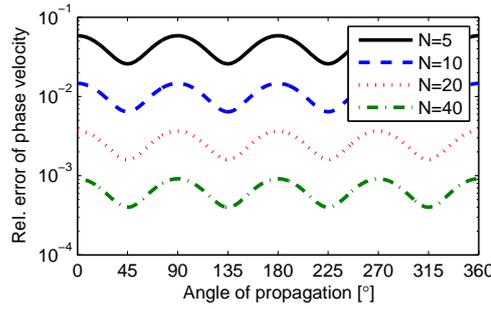
Figure 4.7: *The error in the phase velocity $\tilde{v}$ as a function of the propagation angle in the xy-plane. For all calculations the Courant number was chosen to be $s = 0.577$.*

stay with a finite-difference discretization, there are a variety of proposed techniques [28, 32, 33]. The simplest of which is to increase the order of the spatial discretization by considering more neighboring points. A fourth-order accurate discretization on a staggered grid reads

$$\frac{\partial}{\partial x}F(x_0) = \frac{-F(x_0 + \frac{3\Delta}{2}) + 27F(x_0 + \frac{\Delta}{2}) - 27F(x_0 - \frac{\Delta}{2}) + F(x_0 - \frac{3\Delta}{2})}{24\Delta} + O(\Delta^4)$$

(4.10)

and can be used to derive update equations which are accurate to fourth order.

Here, we will only increase the order of the spatial discretization, while the time integration remains a second-order scheme. The reason is that a higher-order finite-difference scheme in time requires us to store a set of previous field values, thereby dramatically increasing the memory requirements. Should the error be dominated by an insufficient time integration, we can simply reduce Δt to lower this effect. Alternatively, we can use more advanced integration methods as will be discussed in Chapter 6.

By inserting the plane wave ansatz (4.7) into fourth-order update equations results in the numerical dispersion relation

$$\tilde{\omega}_{4\text{th}} = \frac{2}{\Delta t}\arcsin\left(\xi_{4\text{th}}\right),$$

(4.11)

where now

$$\xi_{4\text{th}} = \Delta t\sqrt{\zeta_x^2 + \zeta_y^2 + \zeta_z^2} \quad \text{with} \quad \zeta_i = \frac{27\sin\left(\frac{\tilde{k}_i\Delta i}{2}\right) - \sin\left(\frac{3\tilde{k}_i\Delta i}{2}\right)}{24\Delta i}, \, i = x, y, z.$$

A series expansion of Eq. (4.11) then gives

$$\tilde{\omega} = \sqrt{\tilde{k}_x^2 + \tilde{k}_y^2 + \tilde{k}_z^2} + \frac{1}{24}\left(\tilde{k}_x^2 + \tilde{k}_y^2 + \tilde{k}_z^2\right)^{\frac{3}{2}}\Delta t^2 + O(\Delta^4, \Delta t^4),$$
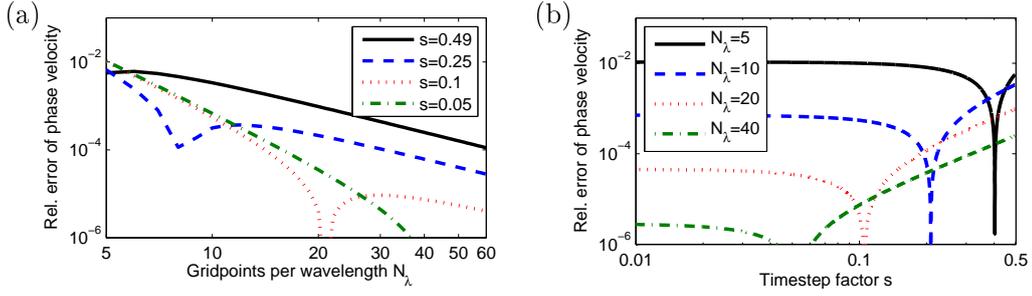
Figure 4.8: *The error in the phase velocity for a fourth-order discretization as a function of (a) the number of gridpoints per wavelength and (b) the timestep factor s. Both are double logarithmic plots and propagation is assumed along the x-axis.*
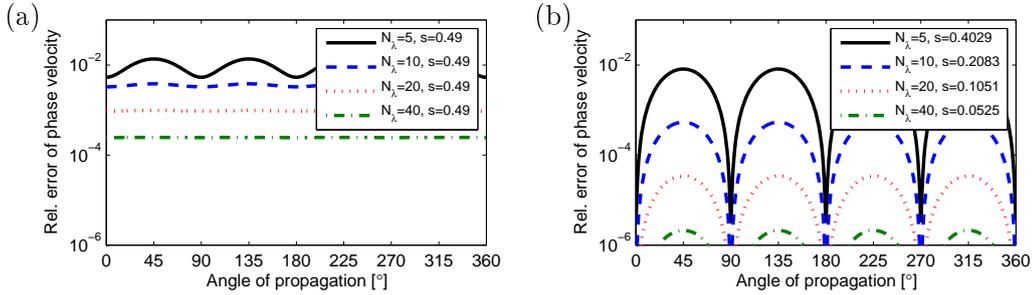
Figure 4.9: *Anisotropy of the phase velocity for a fourth-order discretization as a function of the propagation angle. (a) shows data for the maximal Courant number $s \approx 0.49$, while (b) contains results for optimized s values taken from Fig. 4.8(b).*

from which we recognize the second-order accuracy in time while the error is of $O(\Delta^4)$ in space. Enforcing a purely real $\tilde{\omega}_{4\text{th}}$ results in the new stability criterion

$$\Delta t_{4\text{th}} \leq \frac{6}{7} \frac{1}{\sqrt{\frac{1}{\Delta x^2} + \frac{1}{\Delta y^2} + \frac{1}{\Delta z^2}}},$$

which is reduced by a factor $\frac{6}{7} \approx 0.857$ compared to the second-order formulation. To compare the phase velocity we again assume a cubic grid and propagation in the xy-plane via $\tilde{k}_x = \cos(\varphi)\tilde{k}$, $\tilde{k}_y = \sin(\varphi)\tilde{k}$ and $\tilde{k}_z = 0$. As in the second-order case, we plot the deviation from the exact phase velocity for propagation along the x-axis as a function of N_λ and s in Fig. 4.8. What we can observe directly is a generally reduced error when compared with the data in Fig. 4.6. Even for only five points per wavelength, we now get less than 1% deviation, while the 0.1% mark can be reached with around ten points for an appropriate Δt. Further, we observe certain combinations of s and N, where the error vanishes entirely. To study this behavior further, we look at the anisotropy of the phase velocity error in Fig. 4.9. For a constant $s = 0.49$, we observe how the anisotropy is reduced for larger N_λ. To go a step further, we pick a value of s

which results in a vanishing error for $\varphi = 90°$ (cf. Fig. 4.8(b)), then we also obtain a dramatic reduction in the propagation error for all angles as shown in Fig. 4.9(b).

So far, the implementation of fourth-order discretizations seems like an excellent way to reduce the propagation errors with very little additional effort. One can go even further and obtain tailored finite-difference discretizations which are constructed to minimize this error [28]. Unfortunately, one should keep in mind that we were only dealing with propagation in homogeneous media so far. As soon as we include different materials, two problems occur. First, the optimal value for s depends on the local velocity of light in a medium. So the error can only be minimized for propagation in one particular medium. Second and much more severe is the problem of interfaces. As discussed in the introduction, the electric and magnetic fields are generally not continuous across interfaces. This strongly affects the validity of the Taylor expansion performed to obtain the finite-difference discretizations. Indeed, in Sec. 4.4.2 we will demonstrate how an interface reduces the convergence to second order, despite the higher-order discretizations.

4.3 Making it Work in Practice

Before we can employ the FDTD method in practice, some further issues need to be addressed. In particular, we will briefly discuss the discretization of the material parameters. Further, we also need ways to inject radiation into the system and to formulate appropriate boundary conditions. Finally, we return to the modeling of dispersive materials.

4.3.1 Discretization of Material Parameters

So far, we have discretized the material parameters ϵ and μ at the same location as their corresponding fields. Considering an object with a given ϵ embedded in vacuum, each gridpoint then lies either inside the object or not. This leads to the characteristic stair-casing approximation as sketched in Fig. 4.10. If we now slightly modify the discretization, e.g. by refining it, some of the cells will change their value of ϵ. Thus, even an infinitesimal change in the grid can lead to a significant change of the discretized ϵ-distribution. As a consequence, the convergence of the method is not monotonic. Instead, a refinement might even lead to a significantly higher error. Furthermore, the stair-casing can also lead to a reduction of the convergence to first order [34].

The root of this problem lies in the value of ϵ for a cell which is located on the surface of a discontinuous object. Such a cell contains contributions from multiple materials. A very basic idea to mitigate this problem is to introduce an effective material parameter ϵ_{eff} to fill the cell with. Possibly the simplest idea is to employ a volume average. Indeed, such a procedure leads to a monotonic convergence but it might still have the problem of reduced accuracy. Recently, it was proposed to employ an *anisotropic* effective $\underline{\underline{\epsilon}}_{\text{eff}}$ at a discontinuous interface [35] as

$$\underline{\underline{\epsilon}}_{\text{eff}}^{-1} = \mathcal{P} \left\langle \epsilon^{-1} \right\rangle + \left(1 - \mathcal{P}\right) \left\langle \epsilon \right\rangle^{-1} . \tag{4.12}$$
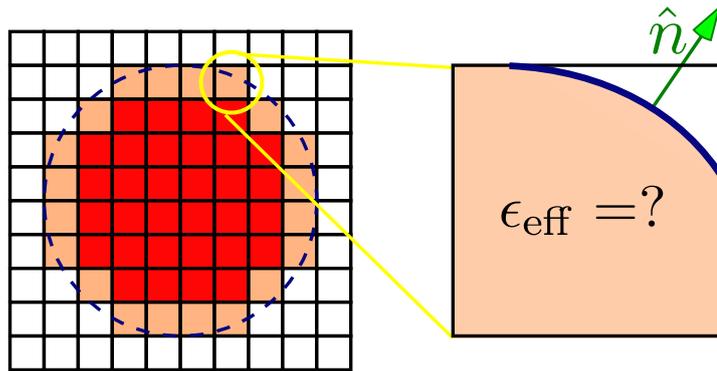
Figure 4.10: *Sketch of a circle, discretized by a uniform mesh. In order to improve the convergence, all cells intersecting with the surface are assigned an effective material parameter.*

Here, $\mathcal{P}$ is the projection matrix $P_{ij} = n_i n_j$ onto the normal vector $\hat{n}$ of the surface. The $\langle \ldots \rangle$ denotes the volume average over one cell. It was demonstrated that this effective material parameter allows to retain second order of accuracy, despite the field discontinuities at the surfaces [35]. Unfortunately, this improvement does not come for free, since it raises the memory requirements and leads to a significantly more involved implementation. Interestingly, for two-dimensional problems in TM polarization, Eq. (4.12) reduces to the mean value, so we can use the scalar parameter $\epsilon_{\text{eff}} = \langle \epsilon \rangle$.

4.3.2 Implementation of Sources

The most obvious way to excite a system is via the current source terms in Maxwell's equations. By simply adding a discretized source term $j|_{i,j,k}$ to the update equations, we implement point-like sources with arbitrary time shapes. Via multiple of such point-sources it is also possible to launch spatially extended pulses, but each point-source will always radiate to all directions, so a planar source will always emit to both sides of the plane.

Total-Field/Scattered-Field Sources

In order to inject an extended pulse travelling only in a given direction into the computational domain, a special technique called total-field/scattered-field (TF/SF) method was developed. The idea relies on the linearity of Maxwell's equations, which allows us to arbitrarily split the total electric and magnetic field into two parts

$$\vec{E}^{(\text{tot})} = \vec{E}^{(\text{inc})} + \vec{E}^{(\text{scat})} \quad \text{and} \quad \vec{H}^{(\text{tot})} = \vec{H}^{(\text{inc})} + \vec{H}^{(\text{scat})}. \tag{4.13}$$

Here, $\vec{E}^{(\text{inc})}$ and $\vec{H}^{(\text{inc})}$ denote the incoming fields. We assume to know these fields analytically, e.g. they are plane waves as given by Eq. (2.17). Furthermore, we split our
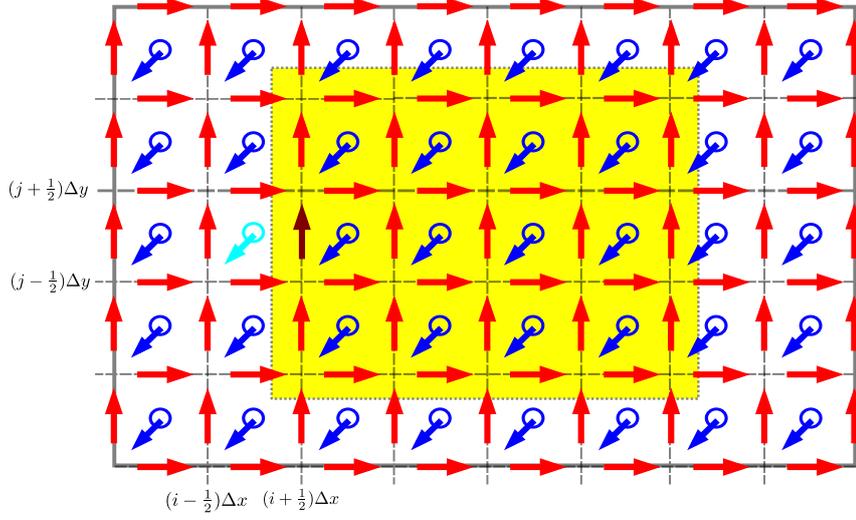
Figure 4.11: *Sketch of a staggered grid in two dimensions (TM polarization) to illustrate the TF/SF method. The shaded region signifies the total field region, while the unshaded areas belong to the scattered field region. The time-stepping of the two marked field components at the boundary is discussed in the text.*

computational domain into two parts. In one region we work with the total fields $\vec{E}^{(\text{tot})}$, $\vec{H}^{(\text{tot})}$ while the other part only contains the scattered fields $\vec{E}^{(\text{scat})}$ and $\vec{H}^{(\text{scat})}$. As an example, we will discuss a two-dimensional system in TM polarization as sketched in Fig. 4.11. Inside each region, the time evolution of a field component only depends on other fields in the same region. At the boundaries however, we see that we need to mix fields from different regions in order to obtain an update equation. As an example, we consider the magnetic field component $H_y^{(\text{tot})}|_{i+\frac{1}{2},j}$ in Fig. 4.11. Its update equation is given by

$$H_y^{(\text{tot})}|_{i+\frac{1}{2},j}^{n+1} = H_y^{(\text{tot})}|_{i,j+\frac{1}{2}}^{n+1} + \frac{1}{\mu|_{i+\frac{1}{2},j}}\frac{\Delta t}{\Delta x}\left(E_z^{(\text{tot})}|_{i+1,j}^{n+\frac{1}{2}} - E_z^{(\text{tot})}|_{i,j}^{n+\frac{1}{2}}\right),$$

where the problem lies in the fact that the component $E_z^{(\text{tot})}|_{i,j}$ does not reside in the total field region anymore. However, by exploiting Eq. (4.13), we can substitute it with available field components as

$$H_y^{(\text{tot})}|_{i+\frac{1}{2},j}^{n+1} = H_y^{(\text{tot})}|_{i,j+\frac{1}{2}}^{n+1} + \frac{1}{\mu|_{i+\frac{1}{2},j}}\frac{\Delta t}{\Delta x}\left(E_z^{(\text{tot})}|_{i+1,j}^{n+\frac{1}{2}} - E_z^{(\text{scat})}|_{i,j}^{n+\frac{1}{2}} - E_z^{(\text{inc})}|_{i,j}^{n+\frac{1}{2}}\right).$$

A similar treatment is required for the updating of $E_z^{(\text{scat})}|_{i,j}$, where the value of $H_y^{(\text{inc})}|_{i+\frac{1}{2},j}$ enters. In this formulation, the known incoming fields essentially act as a

source and inject the desired wave into the total field region. Without any scatterer, the scattered field region will remain zero. Only if the injected wave hits an object inside the total field region, it will scatter a part of the incoming field which can then penetrate the scattered field region. The formulation will prove itself invaluable when calculating scattering cross sections of small objects as discussed in Chapter 8. It should be noted here, that several optimizations have been proposed in order to efficiently obtain the values of the incoming field. For an extensive review of those details, we refer to Ref. [28].

4.3.3 Boundary Conditions and PMLs

Due to limited computational resources, we can usually only model a small region of a realistic experimental system. This leads to boundaries of the computational domain, on which we need to define appropriate boundary conditions.

PEC and PMC Boundary Conditions

Possibly the simplest boundary conditions is to fix the field values on the boundary to a given value. This is called a Dirichlet boundary condition and can be directly implemented in the FDTD scheme. The choice of whether to fix the magnetic or the electric field components depends on how we terminate the staggered grid. In practice, for each boundary we can choose between:

- Perfect Electric Conductor (PEC) boundary conditions, which corresponds to fixing the tangential electric field components to zero.

- Perfect Magnetic Conductor (PMC) boundary conditions, which corresponds to fixing the tangential magnetic field components to zero.

As is well known from electrostatics [20], such a perfect conductor acts as a mirror for electromagnetic radiation. The only difference between the PEC and the PMC condition lies in the phase jump conducted by the reflected wave. In the case of a PEC, the reflected electric field picks up a phase of π when compared to the incoming field, while the magnetic field remains continuous. For the PMC boundary, the role of the fields is swapped.

Discretization of CFS-PMLs

For most applications, the mirroring boundary conditions are not suitable. Instead we require open boundaries, which effectively absorb all outgoing radiation. In Section 3.2 we demonstrated how this can be achieved by perfectly matched layers and we already derived the auxiliary differential equations which need to be solved. Here, we will discuss the implementation of the technically superior CFS-PMLs. For the fully three-dimensional case, this leads to two auxiliary differential equations (3.16) for each

field component, repeated here as

$$\partial_t E_x = \frac{1}{\epsilon}\left(\frac{1}{\kappa_y}\partial_y H_z - \frac{1}{\kappa_z}\partial_z H_y - F_{xy} + F_{xz}\right),$$

$$\partial_t F_{xy} = \frac{\sigma_y}{\kappa_y^2}\partial_y H_z - \left(\alpha_y + \frac{\sigma_y}{\kappa_y}\right)F_{xy},$$

$$\partial_t F_{xz} = \frac{\sigma_z}{\kappa_z^2}\partial_z H_y - \left(\alpha_z + \frac{\sigma_z}{\kappa_z}\right)F_{xz}.$$

Spatially, we can place the auxiliary fields F_{xy} and F_{xz} at the same position as their associated field component E_x. However, if we also choose to collocate them in time, we need to express F_{xy} and F_{xz} at in-between timesteps because they appear on the r.h.s. of the update equation for E_x. We can do so by linear interpolation in time

$$F_{xy}|^{n+\frac{1}{2}} = \frac{F_{xy}|^n + F_{xy}|^{n+1}}{2} + (\Delta t^2). \tag{4.14}$$

Using this expression allows us to execute the time discretization as

$$E_x|^{n+1} = E_x|^n + \frac{\Delta t}{\epsilon}\left(\frac{1}{\kappa_y}\partial_y H_z - \frac{1}{\kappa_z}\partial_z H_y - \frac{F_{xy}|^{n+1} + F_{xy}|^n}{2} + \frac{F_{xz}|^{n+1} + F_{xz}|^n}{2}\right),$$

$$F_{xy}|^{n+1} = F_{xy}|^n + \frac{\sigma_y\Delta t}{\kappa_y^2}\partial_y H_z - \frac{\Delta t}{2}\left(\alpha_y + \frac{\sigma_y}{\kappa_y}\right)\left(F_{xy}|^{n+1} + F_{xy}|^n\right),$$

$$F_{xz}|^{n+1} = F_{xz}|^n + \frac{\sigma_z\Delta t}{\kappa_z^2}\partial_z H_y - \frac{\Delta t}{2}\left(\alpha_z + \sigma_z\kappa_z\right)\left(F_{xz}|^{n+1} + F_{xz}|^n\right),$$

which is not a fully explicit scheme yet, since for the auxiliary fields, values at timestep $n{+}1$ appear on both sides of the equations. However, one can easily solve the equations for the values at timestep $n + 1$ to obtain

$$F_{xy}|^{n+1} = \frac{2 - \Delta t\left(\alpha_y + \frac{\sigma_y}{\kappa_y}\right)}{2 + \Delta t\left(\alpha_y + \frac{\sigma_y}{\kappa_y}\right)}F_{xy}|^n + \frac{2\Delta t\frac{\sigma_y}{\kappa_y^2}}{2 + \Delta t\left(\alpha_y + \frac{\sigma_y}{\kappa_y}\right)}\partial_y H_z|^{n+\frac{1}{2}},$$

$$F_{xz}|^{n+1} = \frac{2 - \Delta t\left(\alpha_z + \frac{\sigma_z}{\kappa_z}\right)}{2 + \Delta t\left(\alpha_z + \frac{\sigma_z}{\kappa_z}\right)}F_{xz}|^n + \frac{2\Delta t\frac{\sigma_z}{\kappa_z^2}}{2 + \Delta t\left(\alpha_z + \frac{\sigma_z}{\kappa_z}\right)}\partial_z H_y|^{n+\frac{1}{2}}.$$

Insertion of the spatial discretization is straight-forward and all other field components can be obtained by permutations of the indices.

Finally, there is also the question of which values should be chosen for the parameters α, κ and σ to achieve optimal performance. Analytically, this choice does not matter, as long as the absorption of the PMLs is sufficiently high. In practice, we will see that the values of the parameters greatly influence the performance. Particularly,

the discontinuous jump from the computational domain into the PMLs can lead to significant reflection due to the discretization. Thus, it is advisable to have an adiabatic transition from the inner domain to the PMLs. In numerical experiments, a polynomial grading of the type

$$\sigma(x) = \left(\frac{x}{d} \right)^m \sigma_{\max}, \tag{4.15a}$$

$$\kappa(x) = 1 + \left(\frac{x}{d} \right)^m (\kappa_{\max} - 1), \tag{4.15b}$$

$$\alpha(x) = \left(\frac{d - x}{d} \right)^{m_a} \alpha_{\max} \tag{4.15c}$$

proved itself valuable. Here, we assumed the surface of the PMLs to be at $x = 0$ and the thickness of the layer is taken as d. Thus, we have to determine a set of five parameters $(m, \sigma_{\max}, \kappa_{\max}, m_a, \alpha_{\max})$. However, for practical considerations it makes sense to rewrite $\sigma_{\max}$ as

$$\sigma_{\max} = \frac{(m + 1)}{2 Y d} R, \tag{4.16}$$

where $Y = \sqrt{\frac{\mu}{\epsilon}}$ is the impedance and d is the thickness of the layer [28]. Now, R takes the role of a rescaled $\sigma_{\max}$ as the free parameter. This rescaling allows to somewhat decouple the influence of m and R and therefore facilitates the search for optimal parameters. We return to the numerical optimization in Sec. 4.4.3.

4.3.4 Dispersive Materials

As discussed in Chapter 3, certain models of dispersive media can be realized via ADEs. As an example, we consider Eq. (3.10) with only the Drude term. To keep the derivation compact, we restrict the discussion to the equations for the x-components, which read

$$\frac{\partial}{\partial t} E_x = \frac{1}{\epsilon_\infty} \left[\frac{\partial}{\partial y} H_z - \frac{\partial}{\partial z} H_y - j_x \right],$$

$$\frac{\partial}{\partial t} j_x = -\gamma_D j_x + \omega_D^2 E_x.$$

From those two equations, we see that the polarization currents $\vec{j}$ should be spatially collocated with their corresponding electric field components. This allows a proper inclusion into the semi-discrete equations for the electric field as

$$\partial_t E_x |_{i,j+\frac{1}{2},k+\frac{1}{2}} = \frac{1}{\epsilon|_{i,j+\frac{1}{2},k+\frac{1}{2}}} \left(\frac{H_z|_{i,j+1,k+\frac{1}{2}} - H_z|_{i,j,k+\frac{1}{2}}}{\Delta y} \right.$$

$$\left. - \frac{H_y|_{i,j+\frac{1}{2},k+1} - H_y|_{i,j+\frac{1}{2},k}}{\Delta z} - j_x|_{i,j+\frac{1}{2},k+\frac{1}{2}} \right).$$

With respect to time, we pick the discretization to coincide with the magnetic fields, which results in the fully discrete update equation for the x-component of the electric field as

$$E_x|_{i,j+\frac{1}{2},k+\frac{1}{2}}^{n+1} = E_x|_{i,j+\frac{1}{2},k+\frac{1}{2}}^{n} + \frac{\Delta t}{\epsilon|_{i,j+\frac{1}{2},k+\frac{1}{2}}} \left(\frac{H_z|_{i,j+1,k+\frac{1}{2}}^{n+\frac{1}{2}} - H_z|_{i,j,k+\frac{1}{2}}^{n+\frac{1}{2}}}{\Delta y} \right.$$

$$\left. - \frac{H_y|_{i,j+\frac{1}{2},k+1}^{n+\frac{1}{2}} - H_y|_{i,j+\frac{1}{2},k}^{n+\frac{1}{2}}}{\Delta z} - j_x|_{i,j+\frac{1}{2},k+\frac{1}{2}}^{n+\frac{1}{2}} \right).$$

Unfortunately, for the update of $\vec{j}$ we have the problem of $\vec{j}$ appearing on the r.h.s. as well. Similarly to Eq. 4.14, we apply an averaging as

$$j_x|_{i,j+\frac{1}{2},k+\frac{1}{2}}^{n} = \frac{j_x|_{i,j+\frac{1}{2},k+\frac{1}{2}}^{n+\frac{1}{2}} + j_x|_{i,j+\frac{1}{2},k+\frac{1}{2}}^{n-\frac{1}{2}}}{2} + O(\Delta t^2)$$

and insert this for the r.h.s. Solving for $j_x|_{i,j+\frac{1}{2},k+\frac{1}{2}}^{n+\frac{1}{2}}$ then results in the update equation

$$j_x|_{i,j+\frac{1}{2},k+\frac{1}{2}}^{n+\frac{1}{2}} = \frac{2 - \gamma_D \Delta t}{2 + \gamma_D \Delta t} j_x|_{i,j+\frac{1}{2},k+\frac{1}{2}}^{n-\frac{1}{2}} + \frac{2\omega_D^2}{2 + \gamma_D \Delta t} E_x|_{i,j+\frac{1}{2},k+\frac{1}{2}}^{n}.$$

Update equations for all other components and for different dispersion models can be derived in an analogous manner. Finally, it should be noted that dispersive media were not considered in the stability analysis conducted in Sec. 4.2.1. A detailed summary of the influence of the ADEs on the error and the stability is provided in Ref. [36].

4.4 Verification of the Method

An absolute must for any implementation of a numerical method is a thorough and comprehensive verification. Ideally, one should base this verification on the comparison with exact analytical results. However, as discussed previously, such analytical results are only available for relatively simple geometries. Therefore, one sometimes also has to test the method against numerical results obtained by different numerical methods.

4.4.1 Empty Metallic Cavities

As a first verification, we consider an empty metallic cavity, which is bounded by perfect magnetic conductors (PMCs). The eigenmodes and the corresponding frequencies for such a system can be obtained analytically [37]. As an example, we pick a two-dimensional system in TM polarization and prepare an eigenmode of the cavity via initial conditions. Then, we let the system evolve for $T = 50\, T_0$, where T_0 is the
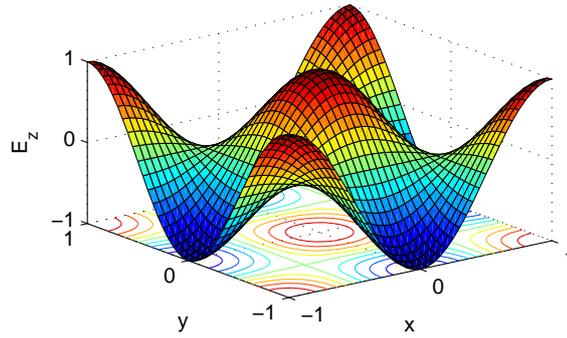
Figure 4.12: *A plot of the cavity mode used for verification. It was picked to have one wavelength in each directions and the corresponding frequency is $\nu = \sqrt{2}$.*

period corresponding to the frequency of the eigenmode. During the calculation, all field values are recorded at all timesteps. After the calculation, we look for the maximum deviation from the analytical solution. The particular mode used for testing has $n_x = n_y = 2$ and is depicted in Fig. 4.12. The corresponding period is $T_0 = 1/\sqrt{2}$. To find the error as a function of time, we record all field values at every timestep. For each timestep, we then compare all recorded values with the exact result and pick the maximum deviation as the error at this given step. This error in time is then normalized to the maximal field value in time and space of the analytic solution. Thus, the relative error in time is defined as

$$\mathcal{E}(t) = \frac{\max\limits_{x,y} \left| E_z^{\text{FDTD}}(x,y,t) - E_z^{\text{exact}}(x,y,t) \right|}{\max\limits_{t} \left(\max\limits_{x,y} \left(E_z^{\text{exact}}(x,y,t) \right) \right)}. \tag{4.17}$$

For the case of the metallic cavity, the denominator of Eq. (4.17) is unity. In Fig. 4.13 we plot a typical behavior of the relative error as a function of time. As expected, one clearly observes a linear growth of the error with time. To finally obtain a measure for the total relative error of a calculation, we again take the maximum of $\mathcal{E}(t)$ to define

$$\mathcal{E}_{\text{tot}} = \max\limits_{t} \left(\mathcal{E}(t) \right). \tag{4.18}$$

The calculations for the two-dimensional metallic cavity are performed with both, second-order and fourth-order spatial discretizations and the relative error in time is plotted as a function of the Courant number and of the discretization in Fig. 4.14. A number of observations can be made from those results. First, we notice that for second-order calculations, a Courant number of $s = s_{\text{max}}$ is optimal which agrees with the theoretical analysis executed in Sec. 4.2.2. For this particular example, the error even seems to vanish entirely for $s = s_{\text{max}}$. In contrast, for fourth-order calculations a sufficiently small time-step is required in order to not be limited by errors from the time integration. Again, there is a distinct minimum which now is resolution dependent
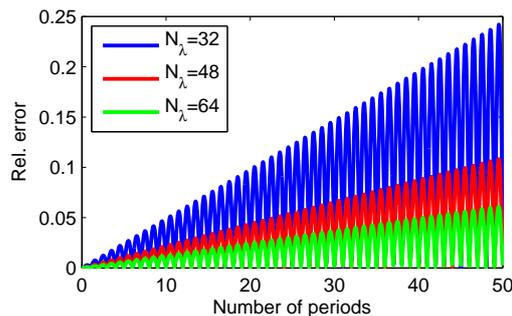
Figure 4.13: *Evolution of the error in time during a FDTD simulation of an empty metallic cavity. Different colors indicate different discretizations. One observes a linear growth of the error with simulation time, where the slope depends on the discretization.*

and does not tend to zero anymore. As we mentioned when studying the theoretical accuracy of the model, purely homogeneous systems are rather academic. Therefore, above calculations only serve to verify our implementation of the method. One should not draw conclusions about the performance and accuracy of the FDTD method based on these results.

4.4.2 Half-Filled Cavities

As a slightly more realistic system which still allows for exact solutions, we next analyze a half-filled cavity. We switch to TE polarization now and the schematic setup is depicted in Fig. 4.15(a). The field component E_x of the used mode is shown in Fig. 4.15(b). In this case, we clearly observe a discontinuity at the interface. In order to avoid problems with the averaging of the material parameters (cf. Sec. 4.3.1), we fix our grid such that the y-components of the electric field are located directly on the interface. For this case, the corresponding ϵ remains scalar and can be averaged to maintain second order accuracy. As we observe from the data in Fig. 4.16, second order convergence is indeed preserved for all cases. However, by looking at the results produced by a fourth-order scheme, we recognize that the convergence is reduced to second order for higher resolutions. This reduction is directly caused by the discontinuities which invalidate the Taylor expansion used to derive Eq. (4.10). Since those errors are spatially localized at the interface, the fourth-order scheme still results in lower total errors than the standard FDTD method.

4.4.3 Optimization of the PML Parameters

To conclude the verification and testing of the FDTD method, we study the performance of the PMLs. As discussed in Sec. 4.3.3, there is a number of free parameters which need to be fixed. As a test system, we use a cube of dimensions $[-1, 1] \times [-1, 1] \times [-1, 1]$ which is discretized with $\Delta x = \Delta y = \Delta z = 0.1$. Outside of
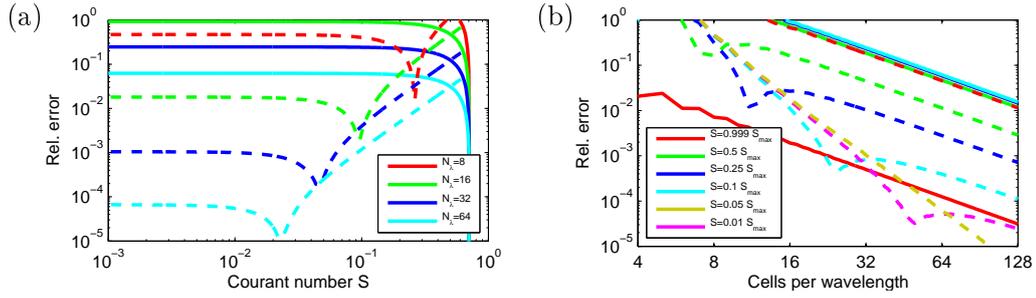
Figure 4.14: *Relative error of a FDTD simulation of a two-dimensional metallic cavity in TM polarization. Solid lines correspond to 2nd-order in space, dashed lines denote a 4th-order discretization. (a) displays the error as a function of the Courant number. The different colors indicate different spatial resolutions. (b) shows the error as a function of N_λ. There, different colors correspond to different Courant numbers.*
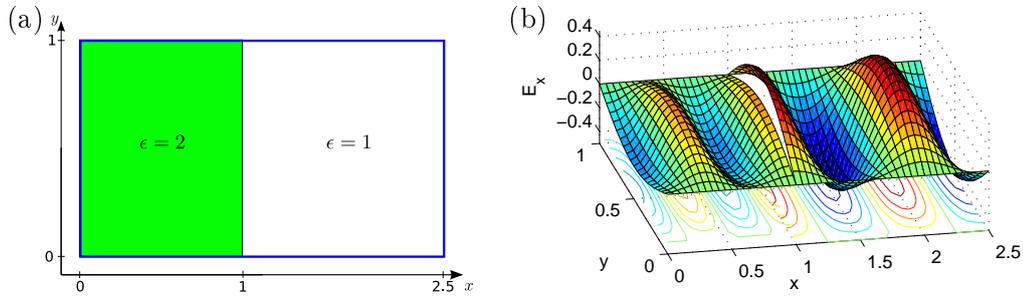
Figure 4.15: *(a) Schematic plot of the half-filled cavity. (b) Plot of the mode used for calculations.*

this cube, we add N_{PML} cells of PMLs, so the thickness of the layers is $d = N_{\mathrm{PML}}\Delta x$. The system is excited via a dipole source at the center of the system. The dipole is oriented along the z-axis and the radiated electric field is recorded at three distinct points $P_1 = (1,0,0)$, $P_2 = (1,1,0)$ and $P_3 = (1,1,1)$. The time variation of the source is given by a Gaussian pulse as

$$ j(t) = \sin\left(2\pi\nu_0(t - t_0)\right)\exp\left(-\frac{(t - t_0)^2}{2w^2}\right). \tag{4.19} $$

We pick the carrier frequency to be $\nu_0 = 1$ and the pulse width $w = 1$. The pulse starts at $t = 0$ and reaches its maximum at $t_0 = 5w$. After $t = 10w$, the source is switched off entirely. In Fig. 4.17 we display the shape of the pulse and the corresponding spectrum. As can be seen from this figure, the pulse has a spectrum which covers wavelength between $\lambda = 8\Delta x$ and $\lambda = 15\Delta x$.

After the calculations, the recorded fields are compared to a reference solution $\vec{E}_{\mathrm{Ref}}$.
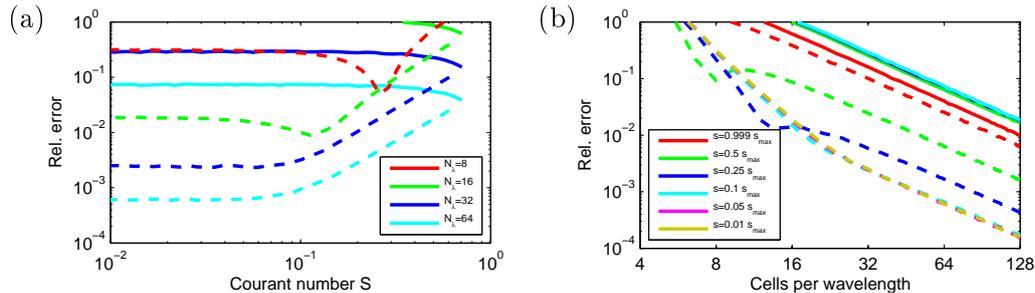
Figure 4.16: *Relative error of a FDTD simulation of a two-dimensional metallic cavity in TM polarization. Solid lines correspond to 2nd-order in space, dashed lines denote a 4th-order discretization. (a) displays the error as a function of the Courant number. The different colors indicate different spatial resolutions. (b) shows the error as a function of N_λ. Here, different colors correspond to different Courant numbers.*
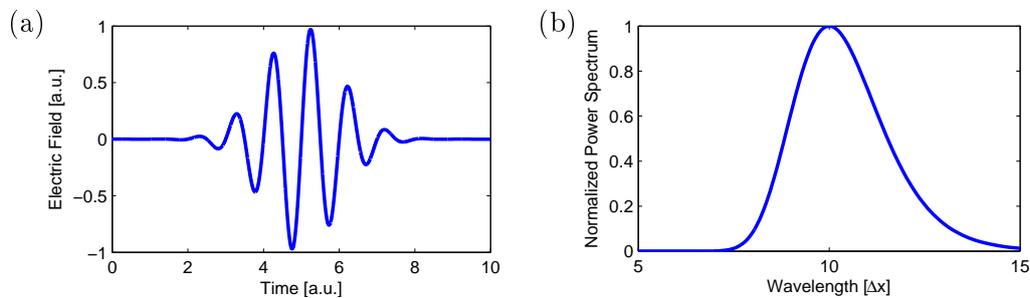
Figure 4.17: *(a) The temporal shape of the pulse used for optimization of the perfectly matched layers. (b) shows the corresponding spectrum.*

The reference solution is obtained by a FDTD calculation of a much larger system but with the same discretization. This procedure allows to exclude propagation errors and to concentrate purely on the artificial reflection of the PMLs. The relative error $\mathcal{E}$ is then calculated as given by Eq. (4.17) and Eq. (4.18).

For the first test, we neglect the parameters α and κ and study only the influence of σ. Thus, we have two free parameters, m and R, to vary. In Fig. 4.18 we present the results of a detailed scan of the two parameters. We directly observe, that the numerical performance does depend strongly on the choice of R and m. This is a purely numerical effect and highlights the importance of the polynomial grading when implementing PMLs in the FDTD method. Concerning the performance, we note a strong dependence on the thickness of the layers. By plotting the minimal error for a given thickness in Fig. 4.19, we observe an exponential reduction with N_{PML}. Even for only five cells, the spurious reflection can be reduced below 1%. For $N_{\mathrm{PML}} = 10$, we already find a large parameter range for which the error is below 0.1%. From these results, we conclude that a PML thickness of around $N_{\mathrm{PML}} = 10$ is advisable. In this
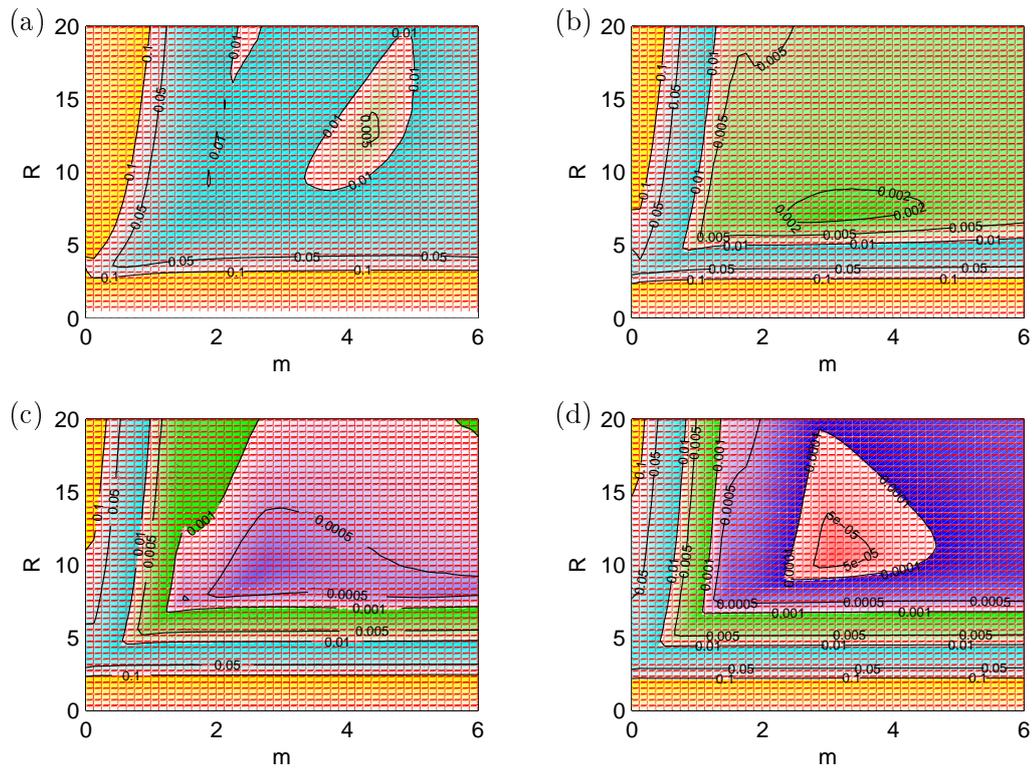
Figure 4.18: *Maps of the maximal error at Point P_3 for different value of R and m. The different panels correspond to a different thickness of the PMLs: (a) $N_{\mathrm{PML}} = 5$, (b) $N_{\mathrm{PML}} = 10$, (c) $N_{\mathrm{PML}} = 15$, (d) $N_{\mathrm{PML}} = 20$.*
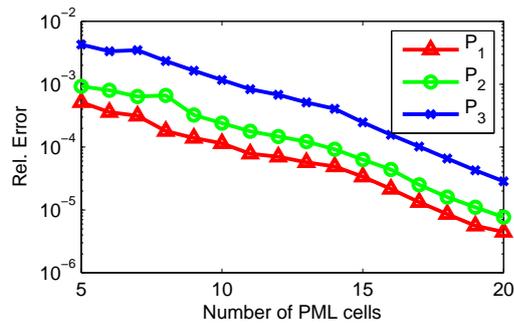
Figure 4.19: *The minimal error obtained as a function of the PML thickness N_{PML}. Here, different symbols indicate the recorded errors at different points in the system.*
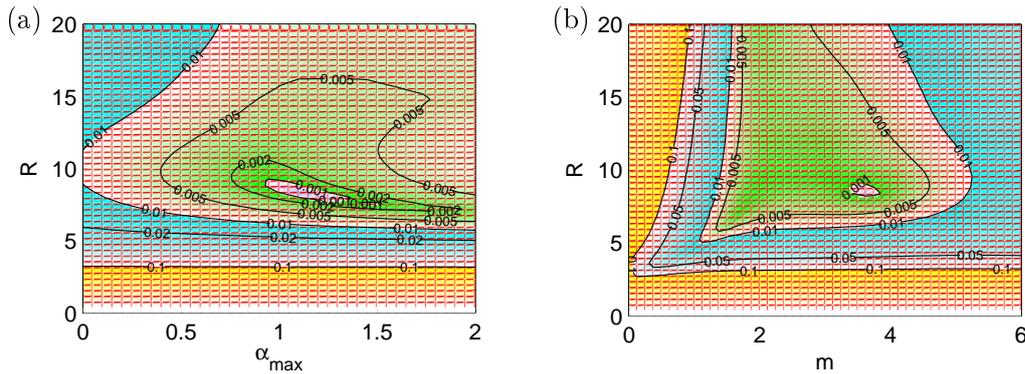
Figure 4.20: *Influence of the PML parameter α on a system with $N_{\mathrm{PML}} = 5$. Here, (a) shows the relative error as a function of R and α_{max} with fixed $m = 3.5$, (b) shows the relative error as a function of R and m with fixed $\alpha_{\mathrm{max}} = 1.0$.*

case, values of $R \approx 10$ and $m \approx 3$ usually lead to a sufficiently good performance.

So far, we only optimized two of the available five parameters in Eqs. (4.15). However, the stretching parameter κ only becomes relevant, if one has a system with evanescent fields leaking into the PMLs. When applied to the absorption of propagating waves, it will usually reduce the performance [27, 28]. Thus, we are left with the free parameter α. In Fig. 4.20 we demonstrate the influence of changing α_{max} on the performance of a system with $N_{\mathrm{PML}} = 5$.

In Fig. 4.20(a), we fixed $m = 3.5$ and plotted the relative error as a function of R and α_{max}. One directly observes an improvement of the performance for finite α. The optimal value for this case seems to be around $\alpha_{\mathrm{max}} = 1$. In Fig. 4.20(b) we also study the influence of non-vanishing α_{max} on the results in the (R, m)-plane. When comparing the data with Fig. 4.18(a), we observe that the improvement does not have a strong influence on the choice of R and m. For a large parameter region, choosing $\alpha_{\mathrm{max}} = 10$ leads to a reduction of the error by roughly a factor 10. Thus, we conclude that a proper optimization of α_{max} can lead to a significant improvement of the absorption. As a consequence, this improvement allows to employ a thinner PML which directly results in faster calculations and reduced memory requirements.

Finally, it should be noted that above test case only covers the absorption of propagating waves. For realistic nanophotonic problems however, evanescent waves often play an important role. If we terminate a system with a scatterer very close to the PMLs, this will lead to significant changes for the choice of optimal PML parameters. As alluded to above, the stretching κ becomes relevant for this case. Unfortunately, there is no general rule on how to select the parameters for optimal absorption of propagating and evanescent waves. A number of numerical experiments can be found in Refs. [27, 28] but the results strongly depend on the particular setup. Thus, one usually has to start from the values proposed above. To verify that the evanescent fields are not distorted by the PMLs, one can shift the boundaries away from the scatterer

until the results stay unchanged. Since evanescent waves decay exponentially, such a procedure is usually feasible and should quickly lead to converging results.

5 Chapter 5

The Discontinuous Galerkin Method

While the FDTD method discussed in the previous chapter is a conceptually simple, yet powerful technique, it has a number of significant shortcomings. The two most important (and somewhat related) problems are the constraint to an orthogonal grid and the limitation to second order when material interfaces are involved. Over the past years, dozens if not hundreds of extensions and alternatives were suggested to overcome some of these restrictions (e.g. see Refs. [28, 38, 39] and references therein). Unfortunately, the price one usually has to pay for a more powerful technique is the loss of simplicity and generality. Hence, none of the proposed discretizations has managed to gain an even remotely similar popularity as the original FDTD method has in the field of nanophotonics.

In this chapter we will discuss a promising alternative technique commonly called discontinuous Galerkin (DG) which attracted some interest recently due to its flexibility and good performance. The method dates back to 1973 where it was originally employed to solve a neutron-transport problem [18]. Since then, there was a steady progress which mostly focussed on problems in hydrodynamics. The major step forward concerning an application to Maxwell's equations was done by Hesthaven and Warburton in 2002 when they proposed a nodal scheme and proved its convergence [19]. In the following sections, a comprehensive review of the DG method with respect to Maxwell's equations will be given. However, the discussion will mainly focus on the parts directly relevant for the implementation. For clarity, some of the mathematical proofs will be omitted and the reader is referred to Refs. [19, 40] for the details.

5.1 One-Dimensional Systems

To initially focus on the fundamentals of the DG method, we start with a one-dimensional system. Repeating Eq. (2.15) without the current term, the electric and
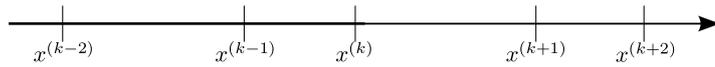
Figure 5.1: *Sketch of a non-uniform tessellation in one dimension.*

magnetic fields are described by

$$\epsilon(x)\frac{\partial}{\partial t}E(x,t) = \frac{\partial}{\partial x}H(x,t), \tag{5.1a}$$

$$\mu(x)\frac{\partial}{\partial t}H(x,t) = \frac{\partial}{\partial x}E(x,t). \tag{5.1b}$$

In order to keep the notation compact, we will restrict the following discussion to the time-evolution of the electric field, i.e., Eq. (5.1a). The corresponding equation for the magnetic field can be obtained later by interchanging the fields and material parameters.

5.1.1 Local mapping of the cells

To discretize Eq. (5.1a) in space, we start by splitting the one-dimensional computational domain into intervals $D^k = [x^k, x^{k+1}]$ as sketched in Fig. 5.1. It is important to note that we don't require a uniform splitting, but the intervals should all have positive length, so $x^{k+1} > x^k$. Next, we introduce a local, linear mapping which maps each cell D^k onto the unit interval $[-1, 1]$. Consequently, a new coordinate u^k in the unit interval is introduced by the transformation

$$u^k(x) = \alpha_0^k + \alpha_1^k x \quad \text{with} \quad \alpha_0^k = -\frac{x^{k+1} + x^k}{x^{k+1} - x^k} \quad \text{and} \quad \alpha_1^k = \frac{2}{x^{k+1} - x^k}.$$

The inverse mapping then reads

$$x(u^k) = \frac{u^k - \alpha_0^k}{\alpha_1^k} = \frac{x^{k+1} + x^k}{2} + \frac{1}{\alpha_1^k}u^k.$$

To be able to transform derivatives and integrals we further require the Jacobian of the transformation, which is given by

$$J^k = \frac{\partial x}{\partial u^k} = \frac{1}{\alpha_1^k}. \tag{5.2}$$

5.1.2 Expansion into a Local Basis

Our next step is to expand the fields in each unit cell into a polynomial basis $\Psi_n(u)$ of order N. Therefore, we need $N_p = N + 1$ coefficients to write

$$E^k(u,t) \approx \sum_{i=1}^{N_p} \hat{E}_i^k(t)\Psi_i(u), \tag{5.3}$$

where $\hat{E}_i^k$ are the expansion coefficients and $\Psi_n(u)$ are arbitrary polynomials of maximal order N. As an example, we could use the so-called monomial basis given by $\Psi_i(u) = u^{i-1}$. In a later section, we will return to the particular choice of the basis. Since we are using a polynomial basis, the same expansion can also be expressed in terms of interpolating Lagrange polynomials $l_i^k(u)$ as

$$E^k(u, t) \approx \sum_{i=1}^{N_p} E^k(u_i, t) l_i^k(u) = \sum_{i=1}^{N_p} E_i^k(t) l_i^k(u). \tag{5.4}$$

Here, we have chosen a set of nodes u_i at which the field is fixed. Between these points, the field values are given by Lagrange interpolation. This representation is also called the *nodal* form, in contrast to the *modal* form of Eq. (5.3). The nodal representation has the advantage that the expansion coefficients are physical field values at given points. This can later be exploited to obtain a better performing implementation of the numerical scheme. Aside from this, the two forms are equivalent.

The next step is to define, in which sense our discrete solution has to agree with the exact solution. A typical way to do so is to pick a set of test functions $\Phi_i^k(u)$ on the unit interval and demand that

$$\int_{D^k} \left(\epsilon(x) \frac{\partial}{\partial t} E(x, t) - \frac{\partial}{\partial x} H(x, t) \right) \Phi_i^k(u^k(x)) dx = 0, \tag{5.5}$$

for all intervals D^k. Similarly to the choice of the expansion basis, there is a certain freedom in choosing the test functions $\Phi_i^k(u)$. A very common and practical choice is to utilize the same set of functions as for the expansion, so $\Phi_i^k(u) = l_i^k(u)$. This choice is known as Galerkin's approach.

Before we proceed with inserting the expansion, it should be noted that the scheme so far is purely local. This means that neighboring cells are not related to each other. To enable this coupling, we employ an integration by parts on Eq. (5.5) to obtain

$$\int_{D^k} \left(\epsilon(x) \frac{\partial}{\partial t} E(x, t) \Phi_i^k(x) + H(x, t) \frac{\partial}{\partial x} \Phi_i^k(x) \right) dx = \left[H(x, t) \Phi_i^k(x) \right]_{x^{(k)}}^{x^{(k+1)}}.$$

This step yields the boundary values on the right-hand side which we now replace by a so-called *numerical flux* $H^*(x, t)$. This procedure might seem slightly unmotivated at this point and we shall return to the details of this substitution later. For now it suffices to say that this numerical flux will facilitate the coupling between neighboring cells, so we have

$$\int_{D^k} \left(\epsilon(x) \frac{\partial}{\partial t} E(x, t) \Phi_i(x) + H(x, t) \frac{\partial}{\partial x} \Phi_i(x) \right) dx = \left[H^*(x, t) \Phi_i(x) \right]_{x^{(k)}}^{x^{(k+1)}}.$$

As a result of the integration by parts, the spatial derivative no longer acts on the fields but on the test functions Φ_i. This form is commonly referred to as the *weak*

formulation. A second integration by parts leads to the equivalent *strong* formulation

$$\int_{D^k} \left(\epsilon(x)\frac{\partial}{\partial t}E(x,t)\Phi_i^k(x) - \frac{\partial}{\partial x}H(x,t)\Phi_i^k(x) \right) dx = \left[\left(H^*(x,t) - H(x,t) \right) \Phi_i^k(x) \right]_{x^{(k)}}^{x^{(k+1)}} .$$

Using this strong formulation, we continue by inserting the expansion (5.4) to obtain

$$\sum_{j=1}^{N} \left(\epsilon^k \mathcal{M}_{ij}^k \frac{\partial}{\partial t}E_j^k(t) - \mathcal{S}_{ij}^k H_j^k(t) \right) = \left[\left(H^*(x,t) - H(x,t) \right) l_i^k(x) \right]_{x^{(k)}}^{x^{(k+1)}}$$

where

$$\mathcal{M}_{ij}^k = \int_{D^k} l_i^k(x) l_j^k(x)\, dx \quad \text{and} \quad \mathcal{S}_{ij}^k = \int_{D^k} l_i^k(x) \frac{\partial}{\partial x} l_j^k(x)\, dx. \tag{5.6}$$

The matrices $\mathcal{M}^k$ and $\mathcal{S}^k$ are called the mass matrix and stiffness matrix, respectively. Here, we already assumed that the material parameters are constant in each cell. Multiplying the system with the inverse mass matrix then results in the final semi-discrete system, which reads

---

Semi-discrete Discontinuous Galerkin Formulation in 1D

$$\epsilon^k \frac{\partial}{\partial t}\mathbf{E}^k(t) = \left(\mathcal{M}^k \right)^{-1} \left(\mathcal{S}^k \mathbf{H}^k(t) + \left[\left(H^*(x,t) - H(x,t) \right) \mathbf{l}^k(x) \right]_{x^{(k)}}^{x^{(k+1)}} \right), \tag{5.7a}$$

$$\mu^k \frac{\partial}{\partial t}\mathbf{H}^k(t) = \left(\mathcal{M}^k \right)^{-1} \left(\mathcal{S}^k \mathbf{E}^k(t) + \left[\left(E^*(x,t) - E(x,t) \right) \mathbf{l}^k(x) \right]_{x^{(k)}}^{x^{(k+1)}} \right). \tag{5.7b}$$

To streamline the notation, we have introduced the vectors $\mathbf{E}^k = \left(E_1^k, \dots, E_{N_p}^k \right)^{\mathrm{T}}$ and $\mathbf{H}^k = \left(H_1^k, \dots, H_{N_p}^k \right)^{\mathrm{T}}$.

5.1.3 The Numerical Flux and the Riemann Problem

The one part still unspecified in our method are the numerical fluxes H^* and E^*, which we combine in a vector $\mathbf{f}^* = \left(H^*, E^* \right)^{\mathrm{T}}$. As alluded to above, the task of the numerical flux is to properly connect neighboring cells. Since our scheme is local, we have two (potentially different) field values at each boundary: One calculated from the expansion in the right cell and one value from the left cell. To obtain a convergent scheme, the numerical flux has to somehow reconcile the two different values. It also must ensure that the physical boundary conditions between two cells are enforced, and finally it needs to guarantee that the resulting method becomes numerically stable.

Fortunately, the idea of a numerical flux is not unique to the DG method. In fact, it is commonly discussed in the context of Finite Volume (FV) methods and a lot of research was conducted towards the proper choice of the numerical flux [41]. While
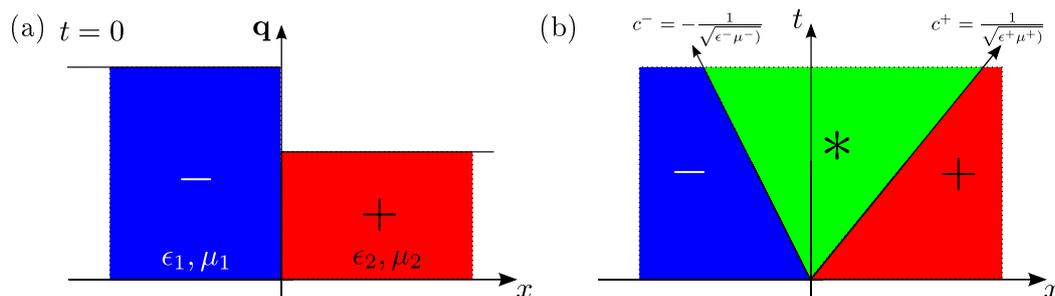
Figure 5.2: *Sketch of a the Riemann problem: (a) shows the initial condition, where the fields have a discontinuity at $x = 0$. (b) displays the evolution of the system in time.*

there is a large variety of possible choices, a particularly suitable numerical flux can be obtained by solving a so-called Riemann problem [41, 42]. The Riemann problem is sketched in Fig. 5.2 and consists of two half-spaces with different, but constant initial fields. Each half-space can also consist of a different material. The question then is, how the fields evolve in time. To find an answer to this question, we return to the full set of equations (5.1) and write them in conservative form as

$$\mathcal{Q}\frac{\partial}{\partial t}\mathbf{q} = \frac{\partial}{\partial x}\mathbf{f}(\mathbf{q}). \tag{5.8}$$

Here, we introduced a vector $\mathbf{q} = (E, H)^T$ and the flux $\mathbf{f}(\mathbf{q}) = (H, E)^T$. The matrix $\mathcal{Q} = \begin{pmatrix} \epsilon(x) & 0 \\ 0 & \mu(x) \end{pmatrix}$ contains the material properties. Multiplication with $\mathcal{Q}^{-1}$ then allows us to write

$$\frac{\partial}{\partial t}\mathbf{q} - \mathcal{B}\frac{\partial}{\partial x}\mathbf{q} = 0$$

with the matrix

$$\mathcal{B} = \begin{pmatrix} \epsilon^{-1} & 0 \\ 0 & \mu^{-1} \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & \epsilon^{-1} \\ \mu^{-1} & 0 \end{pmatrix}.$$

If we diagonalize this matrix as $\mathcal{B} = \mathcal{S}\Lambda\mathcal{S}^{-1}$, we obtain the two eigenvalues $\lambda_{1,2} = \pm\frac{1}{\sqrt{\epsilon\mu}}$. These eigenvalues are the velocities of a forward and a backward travelling wave. For the Riemann problem, on each side only the velocity propagating away from the interface (propagating into the neighboring cell) is relevant. The travelling waves are sketched in Fig. 5.2 by straight lines in the (x, t)-plane, the so-called characteristics. Since the material parameters on each side are different, so are the velocities as indicated by the different slopes of the characteristics.

Our aim is to find a solution for the fields in the central ($*$) region. To do so, we make use of the Rankine-Hugoniot condition [42], which is a consequence of the conservation of $\mathbf{q}$ across the characteristics. For the one-dimensional case, when crossing from region

a to b, the condition reads

$$\boxed{\begin{array}{c} \textsf{Jump Condition} \\[4pt] \forall i: \quad -\lambda_i \mathcal{Q}\left(\mathbf{q}^a - \mathbf{q}^b\right) + \mathbf{f}^a - \mathbf{f}^b = 0. \end{array}} \qquad (5.9)$$

As can be seen in Fig. 5.2, we have two characteristics to cross, first we go from the left $(-)$ to the central $(*)$ region, then from the central $(*)$ to the right $(+)$ region. So, the two conditions read

$$-c^- \mathcal{Q}^- \left(\mathbf{q}^- - \mathbf{q}^*\right) + \mathbf{f}^- - \mathbf{f}^* = 0,$$
$$-c^+ \mathcal{Q}^+ \left(\mathbf{q}^* - \mathbf{q}^+\right) + \mathbf{f}^* - \mathbf{f}^+ = 0,$$

with corresponding velocities $c^- = -\frac{1}{\sqrt{\epsilon^- \mu^-}}$ and $c^+ = \frac{1}{\sqrt{\epsilon^+ \mu^+}}$. Eliminating $\mathbf{q}^*$ and solving the two equations for $\mathbf{f}^*$, we obtain the flux in the $(*)$ region as

$$\mathbf{f}^* = \left[c^+ \mathcal{Q}^+ - c^- \mathcal{Q}^-\right]^{-1} \left[-c^+ c^- \mathcal{Q}^+ \mathcal{Q}^- \left(\mathbf{q}^- - \mathbf{q}^+\right) - c^- \mathcal{Q}^- \mathbf{f}^+ + c^+ \mathcal{Q}^+ \mathbf{f}^-\right].$$

Using $\mathbf{f}^* = \left(H^*, E^*\right)^T$, we can then also derive the fields in the $(*)$ region as

$$E^* = \frac{1}{Y^- + Y^+}\left[Y^- E^- + Y^+ E^+ + \left(H^- - H^+\right)\right], \qquad (5.10a)$$

$$H^* = \frac{1}{Z^- + Z^+}\left[Z^- H^- + Z^+ H^+ + \left(E^- - E^+\right)\right], \qquad (5.10b)$$

with

$$Z^\pm = \sqrt{\frac{\mu^\pm}{\epsilon^\pm}} \quad \text{and} \quad Y^\pm = \left(Z^\pm\right)^{-1} = \sqrt{\frac{\epsilon^\pm}{\mu^\pm}}.$$

Returning to the problem of the numerical flux, it turns out that the solution of the Riemann problem also serves as an excellent numerical flux. This particular choice is also called upwind flux. It should be noted again, that this is by far not the only possible flux. However, in practice this choice leads to a highly stable scheme with little numerical artifacts. The main disadvantage of the flux is, that it does not conserve the total energy in the system. Instead, it leads to dissipation, which can be significant if one does not resolve the propagating waves sufficiently.

5.1.4 The Choice of the Basis and the Nodal Points

To implement the DG scheme, we need to somehow explicitly calculate the matrices $\mathcal{M}^k$ and $\mathcal{S}^k$. To do so, we transform the integrals from Eq. (5.6) onto the unit interval as discussed in Sec. 5.1.1. Then, we have

$$\mathcal{M}_{ij}^k = J^k \int_{-1}^{1} l_i(u) l_j(u)\, du \quad \text{and} \quad \mathcal{S}_{ij}^k = \int_{-1}^{1} l_i(u)\frac{\partial}{\partial u} l_j(u)\, du. \qquad (5.11)$$

N_p	Node Positions u_i	Graphical Representation
3	$-1, 0, 1$	
4	$-1, -\frac{1}{\sqrt{5}}, \frac{1}{\sqrt{5}}, 1$	
5	$-1, -\frac{\sqrt{21}}{7}, 0, \frac{\sqrt{21}}{7}, 1$	
6	$-1, -\sqrt{\frac{7-2\sqrt{7}}{21}}, -\sqrt{\frac{7+2\sqrt{7}}{21}}, \sqrt{\frac{7+2\sqrt{7}}{21}}, \sqrt{\frac{7-2\sqrt{7}}{21}}, 1$	
$\vdots$	$\vdots$	$\vdots$
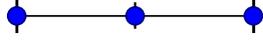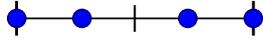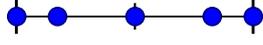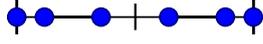
Table 5.1: *Positions of the Legendre-Gauss-Lobatto nodes in the unit interval.*

As a first observation, we note that the matrix $\mathcal{S}^k$ is completely independent of the element k, while the matrix $\mathcal{M}^k$ is almost independent, except for the scalar Jacobian given in Eq. (5.2). This will allow us to precompute these matrices once and reuse them for all cells. It should be noted that this feature is a result of the constant material parameters ϵ and μ in each cell.

The next task concerns the question of how to obtain the Lagrange polynomials $l_i(u)$ and how to integrate over them. In one dimension, the polynomials are given explicitly as

$$l_i(u) = \prod_{j=1, j\neq i}^{N_p} \frac{u - u_j}{u_i - u_j}, \tag{5.12}$$

where u_i are the interpolation nodes. In order to obtain an optimal approximation, one needs to choose these nodes carefully. From the theory of numerical integration, it is known that the zeros of the Legendre polynomials yield the best approximation on the unit interval [43]. However, for implementation reasons that will become clear later, we want two of the nodes fixed at the endpoints of the interval. In this case, the optimal set of interpolation nodes is given by the so-called Legendre-Gauss-Lobatto (LGL) quadrature. The nodal positions up to $N_p = 6$ are presented in Tab. 5.1. Convenient methods to obtain these nodes for arbitrary N_p are discussed in Ref. [40]. For the one-dimensional case, we could proceed by replacing the integrals with a LGL quadrature rule to obtain the desired matrices. While this works very well in one dimension, unfortunately, it can not be generalized easily to higher-dimensional systems. Therefore, we follow Ref. [40] and present a more involved, but also more universal procedure.

Remembering the nodal and modal expansion of E on the unit interval, we have

$$E(u) \approx \sum_{j=1}^{N_p} \hat{E}_j \Psi_j(u) = \sum_{j=1}^{N_p} E_j \, l_j(u) \tag{5.13}$$

with an arbitrary polynomial basis Ψ_i. As alluded to above, we will not use the monomial basis $\Psi_i(u) = u^{i-1}$, because it lacks orthogonality. This might lead to numerical problems later on. Instead, we are free to pick any set of polynomials, which is orthogonal on the unit interval $[-1, 1]$. Here, we choose the normalized Legendre polynomials

$$\tilde{P}_i(u) = \sqrt{i + \frac{1}{2}} P_i(u),$$

where $P_i(u)$ are the classical Legendre polynomials. The orthogonality expresses itself in the property

$$\int\limits_{-1}^{1} \tilde{P}_i(u)\tilde{P}_j(u)\,du = \delta_{ij}. \tag{5.14}$$

From Eq. (5.13), we further recover

$$E(u_i) \approx \sum_{j=1}^{N_p} \hat{E}_j \Psi_j(u_i) = \sum_{j=1}^{N_p} E_j l_j(u_i) = E_i,$$

which allows us to write

$$\mathcal{V}\hat{\mathbf{E}} = \mathbf{E}.$$

Here, we introduced the vectors $\hat{\mathbf{E}} = \left(\hat{E}_1, \ldots, \hat{E}_{N_p}\right)^T$ and $\mathbf{E} = \left(E_1, \ldots, E_{N_p}\right)^T$ as well as the generalized Vandermonde matrix

$$\mathcal{V}_{ij} = \Psi_j(u_i) = \tilde{P}_{j-1}(u_i).$$

Due to the uniqueness of the polynomial interpolation, we also have

$$\tilde{P}_i(u) = \sum_{j=1}^{N_p} \tilde{P}_i(u_j)l_j(u) \Rightarrow \tilde{\mathbf{P}}(u) = \mathcal{V}^T \mathbf{l}(u).$$

and more importantly, the inverse relation

$$l_i(u) = \sum_{j=1}^{N_p} \left(\mathcal{V}^T\right)^{-1}_{ij} \tilde{P}_{j-1}(u). \tag{5.15}$$

Returning to the integrals in Eq. (5.11), we can now exploit Eq. (5.15) and make use of the orthogonality from Eq. (5.14) to express the mass matrix as

$$\mathcal{M}^k_{ij} = J^k \int\limits_{-1}^{1} l_i(u)l_j(u)\,du$$

$$= J^k \sum_{k=1}^{N_p} \sum_{l=1}^{N_p} \left(\mathcal{V}^T\right)^{-1}_{ik} \left(\mathcal{V}^T\right)^{-1}_{jl} \int\limits_{-1}^{1} \tilde{P}_{k-1}(u)\tilde{P}_{l-1}(u)\,du$$

$$= J^k \sum_{k=1}^{N_p} \left(\mathcal{V}^T\right)^{-1}_{ik} \left(\mathcal{V}^T\right)^{-1}_{jk}$$

or more conveniently

$$\mathcal{M}^k = J^k \left(\mathcal{V}\mathcal{V}^T \right)^{-1} = J^k \mathcal{M}. \tag{5.16}$$

In a similar fashion, we can simplify the stiffness matrix $\mathcal{S}$ by rewriting it as

$$\mathcal{S}_{ij} = \int\limits_{-1}^{1} l_i(u) l_j'(u)\, du = \int\limits_{-1}^{1} l_i(u) \sum_{k=1}^{N_p} l_j'(u_k) l_k(u)\, du = \sum_{k=1}^{N_p} \mathcal{M}_{ik} \mathcal{D}_{kj}.$$

Here, the prime denotes the derivative with respect to the argument and we introduced the differentiation matrix

$$\mathcal{D}_{ij} = l_j'(u_i).$$

Making use of the Eq. (5.15), we insert the generalized Vandermonde matrix and obtain

$$\mathcal{D}_{ij} = \sum_{k=1}^{N_p} \left(\mathcal{V}^T \right)_{jk}^{-1} \tilde{P}'_{k-1}(r_i) = \sum_{k=1}^{N_p} \tilde{P}'_{k-1}(u_i) \mathcal{V}_{kj}^{-1}. \tag{5.17}$$

Exploiting an identity for the derivatives of the Legendre polynomials [40] allows us to express $\tilde{P}'_i(u)$ as

$$\tilde{P}'_i(u) = \sqrt{i(i+1)} \tilde{P}_{i-1}^{(1,1)}(u),$$

where $\tilde{P}_i^{(1,1)}$ is a normalized Jacobi polynomial. For details on how to evaluate Jacobi polynomials see Ref. [40].

5.2 Two- and Three-Dimensional Systems

The most obvious problem in the transition from one-dimensional problems to higher dimensions concerns the spatial tessellation. To be able to accurately model complex geometries, we require unstructured meshes. However, the exact shape of the elements still offers a variety of options. Some of the commonly used cell types are shown in Fig. 5.3. We will restrict the discussion to triangular (2D) and tetrahedral (3D) elements, but this is by no means a limitation of the method. We will also skip the topic of mesh generation, because it would go far beyond the scope of this thesis. The interested reader is referred to Ref. [44]. All meshes employed in the following sections were either created by hand or they were generated by using the freely available software NETGEN [44].

To derive a DG scheme for Maxwell's equations in higher dimensions, we start the conservation form (2.11), which we recast to

$$\partial_t \left(\mathcal{Q}(\vec{r}) \mathbf{q}(\vec{r}, t) \right) + \nabla \cdot \vec{F}(\mathbf{q}) = 0. \tag{5.18}$$

The material matrix $\mathcal{Q}(\vec{r})$, the state vector $\mathbf{q}$ and the flux vector $\vec{F}(\mathbf{q}) = \left(\mathbf{F}_x, \mathbf{F}_y, \mathbf{F}_z \right)^{\mathrm{T}}$ are defined as

$$\mathcal{Q}(\vec{r}) = \begin{pmatrix} \epsilon(\vec{r}) & 0 \\ 0 & \mu(\vec{r}) \end{pmatrix}, \ \mathbf{q}(\vec{r}, t) = \begin{pmatrix} \vec{E}(\vec{r}, t) \\ \vec{H}(\vec{r}, t) \end{pmatrix}, \ \text{and } \mathbf{F}_i(\mathbf{q}) = \begin{pmatrix} -\hat{e}_i \times \vec{H}(\vec{r}, t) \\ \hat{e}_i \times \vec{E}(\vec{r}, t) \end{pmatrix}.$$

2d

Triangle   Quadrilateral

3d

Tedrahedron   Hexahedron   Prism   Pyramide
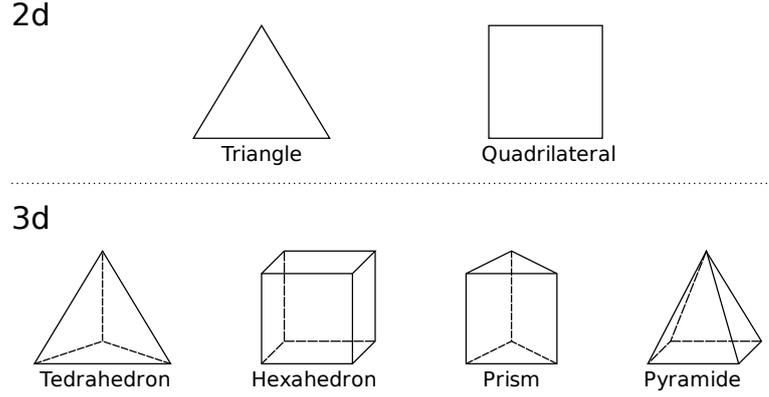
Figure 5.3: *Different common types of cells in two-dimensional (upper panel) and three-dimensional (lower panel) unstructured meshes.*

In these expressions, $\hat{e}_i$, $i = x, y, z$ denote the Cartesian unit vectors. In order to solve this system, the computational domain is tessellated into K conforming elements Ω^k. As discussed above, those elements cyn be of arbitrary shape. Now, the general procedure closely follows the sequence discussed for the one-dimensional case. On each element, the fields are expanded in terms of interpolating Lagrange polynomials $L_i(\vec{r})$ as

$$\mathbf{q}^k(\vec{r}, t) \approx \sum_{i=1}^{N_p} \mathbf{q}^k(\vec{r}_i, t) L_i(\vec{r}) = \sum_{i=1}^{N_p} \tilde{\mathbf{q}}_i^k(t) L_i(\vec{r}), \tag{5.19}$$

where N_p denotes the number of coefficients utilized. The vector $\tilde{\mathbf{q}}^k(t)$ contains the unknown field values that have to be solved for, while $\vec{r}_i$ denotes a set of suitable interpolation nodes. The question of how to obtain those nodes is deferred to a later section.

Application of the standard Galerkin approach then consists of multiplying Eq. (5.18) with test functions $L_i(\vec{r})$ and integrating over an element Ω^k, which results in

$$\int_{\Omega^k} \left(\mathcal{Q}^k \partial_t \tilde{\mathbf{q}}^k + \nabla \cdot \vec{F}(\tilde{\mathbf{q}}^k) \right) L_i(\vec{r}) d\vec{r} = 0.$$

To facilitate the coupling between neighboring cells, we again employ an integration by parts and substitute the physical flux $\vec{F}(\tilde{\mathbf{q}}^k(t))$ with the numerical flux $\vec{F}^*(\tilde{\mathbf{q}}^k(t))$ in the resulting contour integral. A second integration by parts then results in the strong formulation

$$\int_{\Omega^k} \left(\mathcal{Q}^k \partial_t \tilde{\mathbf{q}}^k + \nabla \cdot \vec{F}(\tilde{\mathbf{q}}^k) \right) L_i(\vec{r}) d\vec{r} = \int_{\partial\Omega^k} \hat{n} \cdot \left(\vec{F}(\tilde{\mathbf{q}}^k) - \vec{F}^*(\tilde{\mathbf{q}}^k) \right) L_i(\vec{r}) d\vec{r}, \tag{5.20}$$

where $\hat{n}$ is the outward-pointing normal vector of the cell surface $\partial\Omega^k$.
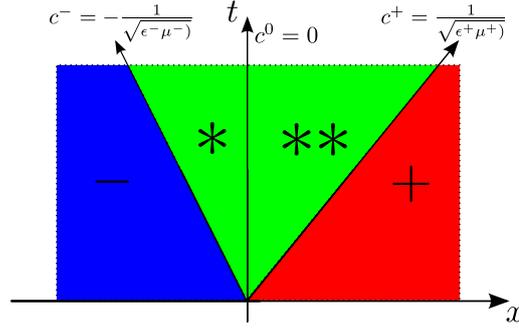
Figure 5.4: *Sketch of a the Riemann problem for Maxwell's equations in two and three dimensions.*

5.2.1 The Numerical Flux in Higher Dimensions

To continue with the derivation of the higher-dimensional equations, we revisit to the topic of the numerical flux. Similarly to the one-dimensional case, we calculate the numerical flux via solving a Riemann problem. We start from the conservative form given in Eq. (2.4), repeated here as

$$\mathcal{Q}\partial_t \mathbf{q} + \nabla \cdot \vec{F} = \mathcal{Q}\partial_t \mathbf{q} + \mathcal{A}_x \partial_x \mathbf{q} + \mathcal{A}_y \partial_y \mathbf{q} + \mathcal{A}_z \partial_z \mathbf{q} = 0. \tag{5.21}$$

Since we are interested in the flux along a normal vector $\hat{n} \cdot \vec{F}$, we consider the equation

$$\mathcal{Q}\partial_t \mathbf{q} + \partial_{\hat{n}} \mathbf{F}_{\hat{n}} = 0,$$

where $\partial_{\hat{n}}$ denotes the directional derivative along $\hat{n}$ and

$$\mathbf{F}_{\hat{n}} = \hat{n} \cdot \vec{F} = \begin{pmatrix} \vec{F}_E \\ \vec{F}_H \end{pmatrix} = \underbrace{\left(n_x \mathcal{A}_x + n_y \mathcal{A}_y + n_z \mathcal{A}_z \right)}_{=:\Pi} \mathbf{q} = \begin{pmatrix} -\hat{n} \times \vec{H} \\ \hat{n} \times \vec{E} \end{pmatrix}.$$

Thereby, we have reduced the fully three-dimensional problem to a one-dimensional problem along the $\hat{n}$-axis and can apply the Rankine-Hugoniot condition (5.9). The eigenvalues of the matrix Π are given by $c^{\pm} = \pm \frac{1}{\sqrt{\epsilon\mu}}$ and $c^0 = 0$, where each eigenvalue is doubly degenerate. The main difference to the one-dimensional case is the third eigenvalue c^0, which leads to a total of four regions in the (x, t)-plane as sketched in Fig. 5.4. In total, the Rankine-Hugoniot condition yields the following three equations

$$-c^- \mathcal{Q}^- \left(\mathbf{q}^- - \mathbf{q}^* \right) + \mathbf{F}_{\hat{n}}^- - \mathbf{F}_{\hat{n}}^* = 0,$$

$$\mathbf{F}_{\hat{n}}^* - \mathbf{F}_{\hat{n}}^{**} = 0,$$

$$-c^+ \mathcal{Q}^+ \left(\mathbf{q}^+ - \mathbf{q}^{**} \right) + \mathbf{F}_{\hat{n}}^+ - \mathbf{F}_{\hat{n}}^{**} = 0.$$

Inserting the explicit expressions for $\mathcal{Q}$, $\mathbf{q}$ and $\mathbf{F}_{\hat{n}}$, allows us to eliminate $\mathbf{q}^*$ and $\mathbf{q}^{**}$ and to solve for $\mathbf{F}_{\hat{n}}^*$. The resulting numerical flux reads

$$
\mathbf{F}_{\hat{n}}^* = \begin{pmatrix} -\hat{n} \times H^* \\ \hat{n} \times E^* \end{pmatrix} = \begin{pmatrix} \frac{1}{Z^+ + Z^-} \hat{n} \times \left[-Z^- \vec{H}^- - Z^+ \vec{H}^+ + \hat{n} \times \Delta \vec{E} \right] \\ \frac{1}{Y^+ + Y^-} \hat{n} \times \left[Y^- \vec{E}^- + Y^+ \vec{E}^+ + \hat{n} \times \Delta \vec{H} \right] \end{pmatrix}.
$$

Here, we have introduced the difference of the fields at the interface as

$$
\Delta \vec{E} = \vec{E}^- - \vec{E}^+ \quad \text{and} \quad \Delta \vec{H} = \vec{H}^- - \vec{H}^+.
$$

As in the one-dimensional case, the superscript "-" denotes the local cell while "+" signifies the neighboring element.

For the strong formulation as given in (5.20), we require the difference between the physical and the numerical flux. This difference can be expressed as

$$
\boxed{
\begin{array}{c}
\text{Numerical Flux for the Strong Formulation} \\
\hat{n} \cdot \left(\vec{F} - \vec{F}^* \right) = \begin{pmatrix} \vec{F}_E^- - \vec{F}_E^* \\ \vec{F}_H^- - \vec{F}_H^* \end{pmatrix} = \begin{pmatrix} \hat{n} \times \frac{Z^+ \Delta \vec{H} - \hat{n} \times \Delta \vec{E}}{Z^- + Z^+} \\ \hat{n} \times \frac{-Y^+ \Delta \vec{E} - \hat{n} \times \Delta \vec{H}}{Y^- + Y^+} \end{pmatrix}.
\end{array}
} \tag{5.22}
$$

5.2.2 Briefly on Boundary Conditions

In addition to the equations and an initial condition, every linear PDE must also be complemented by appropriate boundary conditions. In our case, a number of different conditions are possible. Typical choices are either reflecting or open boundaries.

Reflective Boundary Conditions

As discussed in Sec. 4.3.3, we can differentiate between two types of reflective boundary conditions, namely either perfect electric conductors (PECs) or perfect magnetic conductors (PMCs). In case of PECs, we enforce that the tangential component of the electric field vanishes at the boundary. For PMCs, the tangential components of the magnetic field have to be zero at the surface. Both PEC and PMC are reflective with a phase shift of π acquired by either the electric or the magnetic field, respectively. Explicitly, for the PEC case the boundary condition reads

$$
\hat{n} \times \vec{E} = 0
$$

with no particular constraint on $\vec{H}$. This can be implemented by exploiting the mirror principle as

$$
\hat{n} \times \vec{E}^+ = -\hat{n} \times \vec{E}^-.
$$

Additionally, for the magnetic field we simply choose $\hat{n} \times \vec{H}^+ = \hat{n} \times \vec{H}^-$. Those conditions can be easily integrated into the flux calculation, by modifying $\Delta \vec{E}$ and $\Delta \vec{H}$ on the outer boundary. Then, we also have to set $Y^+ = Y^-$ and $Z^+ = Z^-$. The particular values for $\Delta \vec{E}$ and $\Delta \vec{H}$ are summarized in Tab. 5.2.

	$\mathbf{\Delta \tilde{E}}$	$\mathbf{\Delta \tilde{H}}$
Perfect Electric Conductor (PEC)	$-2\vec{E}^-$	0
Perfect Magnetic Conductor (PMC)	0	$-2\vec{H}^-$
Silver-Müller (1st order absorbing)	$-2\vec{E}^-$	$-2\vec{H}^-$

Table 5.2: *Implementation of different boundary conditions by modifying $\Delta \vec{E}$ and $\Delta \vec{H}$.*

Silver-Müller Boundary Conditions

For many realistic problems, it would be advantageous to have a simple way to implement open boundary conditions. Open in this context means that all radiation hitting the outer boundaries leaves the system and never returns. Analytically, one can impose the condition of out-going radiation via the Silver-Müller radiation conditions [45]

$$\lim_{r \to \infty} r \left((\nabla \times \vec{E}) \times \hat{r} - ik\vec{E} \right) = 0,$$
$$\lim_{r \to \infty} r \left((\nabla \times \vec{H}) \times \hat{r} - ik\vec{H} \right) = 0.$$

Here, $r = |\vec{r}|$ is the distance to a scatterer, $\hat{r} = \vec{r}/r$ signifies the direction of propagation and $k = \omega$ (in dimensionless units) denotes the wave vector. Those expressions suggest, that one could try to obtain approximate open boundary conditions at finite r by enforcing

$$(\nabla \times \vec{E}) \times \hat{n} - ik\vec{E} = 0,$$
$$(\nabla \times \vec{H}) \times \hat{n} - ik\vec{H} = 0,$$

at the outer boundaries with outward pointing normal vector $\hat{n}$. Inserting the curl equations in frequency domain, we obtain

$$i\omega \vec{H} \times \hat{n} - ik\vec{E} = 0,$$
$$-i\omega \vec{E} \times \hat{n} - ik\vec{H} = 0.$$

After some rearranging and another cross-product with $\hat{n}$, we finally recover the boundary conditions

$$-\hat{n} \times \vec{H} = -\hat{n} \times \hat{n} \times \vec{E},$$
$$\hat{n} \times \vec{E} = -\hat{n} \times \hat{n} \times \vec{H}.$$

Those expressions are called Silver-Müller boundary conditions and to enforce them, we can simply exploit the numerical flux by requiring

$$\mathbf{F}_{\hat{n}}^* = \begin{pmatrix} -\hat{n} \times H^* \\ \hat{n} \times E^* \end{pmatrix} = \begin{pmatrix} -\hat{n} \times \hat{n} \times \vec{E}^- \\ -\hat{n} \times \hat{n} \times \vec{H}^- \end{pmatrix}.$$

As a result, we derive

$$\hat{n} \cdot \left(\vec{F} - \vec{F}^* \right) = \begin{pmatrix} \vec{F}_E^- - \vec{F}_E^* \\ \vec{F}_H^- - \vec{F}_H^* \end{pmatrix} = \begin{pmatrix} -\hat{n} \times H^- + \hat{n} \times \hat{n} \times \vec{E}^- \\ \hat{n} \times E^- + \hat{n} \times \hat{n} \times \vec{H}^- \end{pmatrix}. \qquad (5.23)$$

Similarly to the case of reflective boundary conditions, we can implement this case by setting $\Delta \vec{E}$ and $\Delta \vec{H}$ as given in Tab. 5.2. It should be noted that the Silver-Müller boundary conditions are only approximate and they improve as the distance to the boundaries tends to infinity. For most practical applications, they are *not sufficient* and still result in rather strong reflections from the boundaries. In those cases, we need to employ PMLs, which will be discussed in Sec. 5.4.

5.2.3 Derivation of the Semi-Discrete System

With the flux and the boundary conditions in place, we can continue to derive a semi-discrete system. Inserting the numerical flux (5.22) into Eq. (5.20) leads us to the expressions

$$\int_{\Omega^k} \left(\epsilon^k \partial_t \vec{E}^k - \nabla \times \vec{H} \right) L_i(\vec{r}) d\vec{r} = \int_{\partial \Omega^k} \left(\hat{n} \times \frac{Z^+ \Delta \vec{H} - \hat{n} \times \Delta \vec{E}}{Z^- + Z^+} \right) L_i(\vec{r}) d\vec{r},$$

$$\int_{\Omega^k} \left(\mu^k \partial_t \vec{H}^k + \nabla \times \vec{E} \right) L_i(\vec{r}) d\vec{r} = \int_{\partial \Omega^k} \left(\hat{n} \times \frac{-Y^+ \Delta \vec{E} - \hat{n} \times \Delta \vec{H}}{Y^- + Y^+} \right) L_i(\vec{r}) d\vec{r}.$$

Now, we can insert the expansion of the fields (5.19) to obtain

$$\epsilon^k \mathcal{M}^k \partial_t \tilde{\vec{E}}^k = \vec{\mathcal{S}}^k \times \tilde{\vec{H}}^k + \mathcal{F}^k \left(\hat{n} \times \frac{Z^+ \Delta \tilde{\vec{H}}^k - \hat{n} \times \Delta \tilde{\vec{E}}}{Z^- + Z^+} \right), \qquad (5.24a)$$

$$\mu^k \mathcal{M}^k \partial_t \tilde{\vec{H}}^k = -\vec{\mathcal{S}}^k \times \tilde{\vec{E}} + \mathcal{F}^k \left(\hat{n} \times \frac{-Y^+ \Delta \tilde{\vec{E}} - \hat{n} \times \Delta \tilde{\vec{H}}}{Y^- + Y^+} \right), \qquad (5.24b)$$

where we introduced the new notation

$$\tilde{\vec{E}}^k = \left(\tilde{\mathbf{E}}_x^k, \tilde{\mathbf{E}}_y^k, \tilde{\mathbf{E}}_z^k \right) \quad \text{and} \quad \tilde{\vec{H}}^k = \left(\tilde{\mathbf{H}}_x^k, \tilde{\mathbf{H}}_y^k, \tilde{\mathbf{H}}_z^k \right)$$

Further, we used a vector $\vec{\mathcal{S}}^k = \left(\mathcal{S}_x^k, \mathcal{S}_y^k, \mathcal{S}_z^k \right)$ and the matrices

$$\left(\mathcal{S}_m^k \right)_{ij} = \int_{\Omega^k} L_i(\vec{r}) \partial_m L_j(\vec{r}) d\vec{r}, \text{ with } m \in \{x, y, z\}, \qquad (5.25a)$$

$$\left(\mathcal{M}^k \right)_{ij} = \int_{\Omega^k} L_i(\vec{r}) L_j(\vec{r}) d\vec{r}, \qquad (5.25b)$$

$$\left(\mathcal{F}^k \right)_{ij} = \int_{\partial \Omega^k} L_i(\vec{r}) L_j(\vec{r}) d\vec{r} \text{ where } j \in \{j | \vec{r}_j \in \partial \Omega^k\}. \qquad (5.25c)$$

As in the one-dimensional case, $\mathcal{M}$ and $\mathcal{S}$ are the mass matrix and the stiffness matrix, respectively. Additionally, we introduced a new matrix $\mathcal{F}^k$, which only acts on the surface of a cell. Depending on the dimensionality of the system, we call this matrix edge-mass (2D) or face-mass matrix (3D).

5.2.4 Mapping of the Cells and Node Generation

As in the one-dimensional case, we would like to perform as many operations as possible on a unit cell. To be able to do so, we need to find a transformation from an arbitrary cell to our reference cell.

Two Dimensions

In the two-dimensional case, this unit cell is chosen to be a right triangle as depicted in Fig. 5.5. As alluded to above, we require a linear transformation, which leads us to
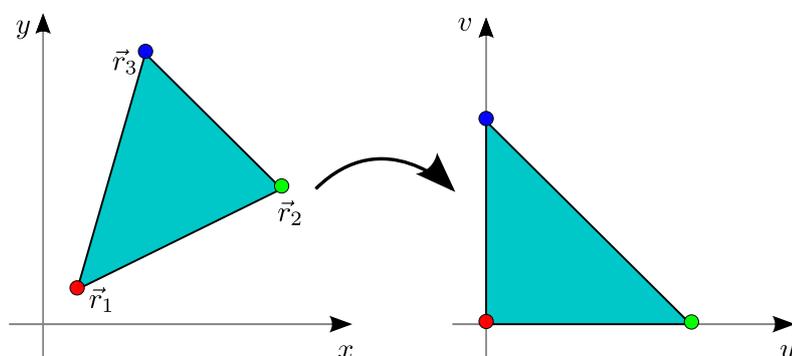
Figure 5.5: *A sketch of the two-dimensional mapping onto a reference triangle (unit cell).*

the general two-dimensional mapping

$$x = x_0 + \mathcal{J}_{11}^{(2d)} u + \mathcal{J}_{12}^{(2d)} v, \tag{5.26a}$$

$$y = y_0 + \mathcal{J}_{21}^{(2d)} u + \mathcal{J}_{22}^{(2d)} v. \tag{5.26b}$$

In the case of an arbitrary triangle with vertices $\vec{r}_1 = (x_1, y_1)$, $\vec{r}_2 = (x_2, y_2)$ and $\vec{r}_3 = (x_3, y_3)$, the transformation coefficients are uniquely determined by solving the system (5.26) for each vertex. The results then read

$$x_0 = x_1, \qquad\qquad y_0 = y_1,$$
$$\mathcal{J}_{11}^{(2d)} = x_2 - x_1, \qquad\qquad \mathcal{J}_{21}^{(2d)} = y_2 - y_1,$$
$$\mathcal{J}_{12}^{(2d)} = x_3 - x_1, \qquad\qquad \mathcal{J}_{22}^{(2d)} = y_3 - y_1.$$

Here, the matrix $\mathcal{J}^{(2d)}$ is the metric of the transformation and also called Jacobi matrix. The Jacobian $J^{(2d)} = \det\left(\mathcal{J}^{(2d)}\right)$ is then given by

$$J^{(2d)} = (x_2 - x_1)(y_3 - y_1) - (x_3 - x_1)(y_2 - y_1).$$

In the two-dimensional case, the expansion order N and the number of coefficients N_p is related via

$$N_p = \frac{(N+1)(N+2)}{2}.$$

Just as in the one-dimensional case, we require an orthogonal basis on the reference triangle, to avoid a poorly conditioned Vandermonde matrix. Fortunately, such a basis is well known and is often referred to as Koornwinder-Dubiner polynomials. On the unit cell shown in Fig. 5.5, they take the form

$$\Psi_m = \frac{\sqrt{8(2i+1)(i+j+1)}}{(2i+2)} P_i^{(0,0)}(a) P_j^{(2i+1,0)}(b)(1-b)^i$$

with

$$a = \frac{2u}{1-v} - 1, \quad \text{and} \quad b = 2v - 1.$$

In Fig. 5.6, we have plotted the first few Koornwinder-Dubiner polynomials, to give an impression of how the basis looks.

The final missing ingredient is the set of nodes $\vec{r}_i$ on the triangle. Unfortunately, there is no known way to analytically obtain an optimal set in higher dimensions. However, there is a number of proposals how to obtain reasonably good interpolation nodes on triangles and tetrahedra [46–48]. We continue by following the proposal of Warburton as given in Ref. [47]. This method is called "warp & blend" and it enforces the nodal points on the edges to be LGL points. We will not repeat the algorithm here, but instead simply show the resulting positions for several expansion orders in Fig. 5.7.

Three Dimensions

In three dimensions, we choose the unit cell to be an equilateral tetrahedron as displayed in Fig. 5.8. The general procedure is now almost identical to the two-dimensional case. Assuming a general linear mapping of the form

$$\vec{r} = \vec{r}_0 + \mathcal{J}^{(3d)} \begin{pmatrix} u \\ v \\ w \end{pmatrix} \tag{5.27}$$

directly yields

$$\vec{r}_0 = \frac{\vec{r}_1 + \vec{r}_2}{2}.$$
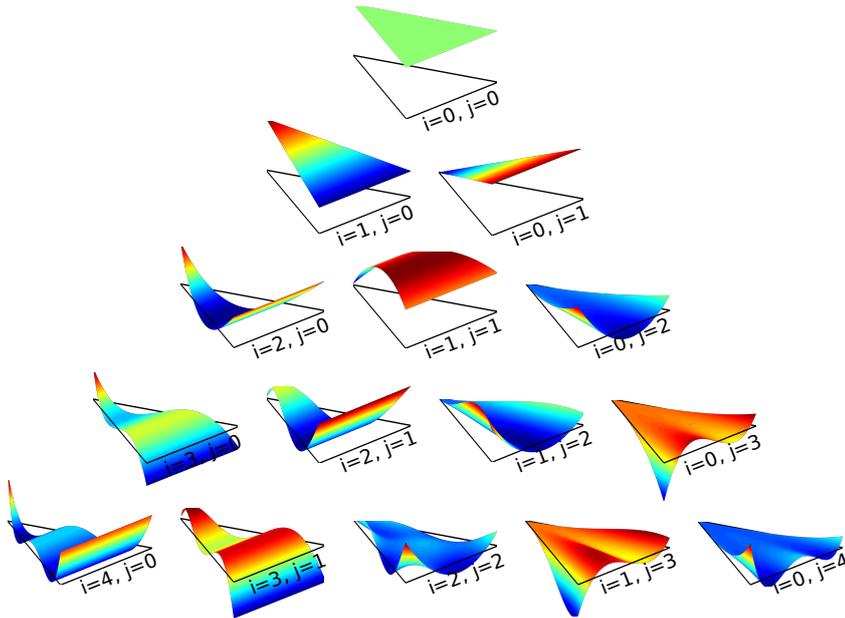
Figure 5.6: *Plot of Koornwinder-Dubiner polynomials. Each row corresponds to the additional basis functions required by a higher expansion order.*
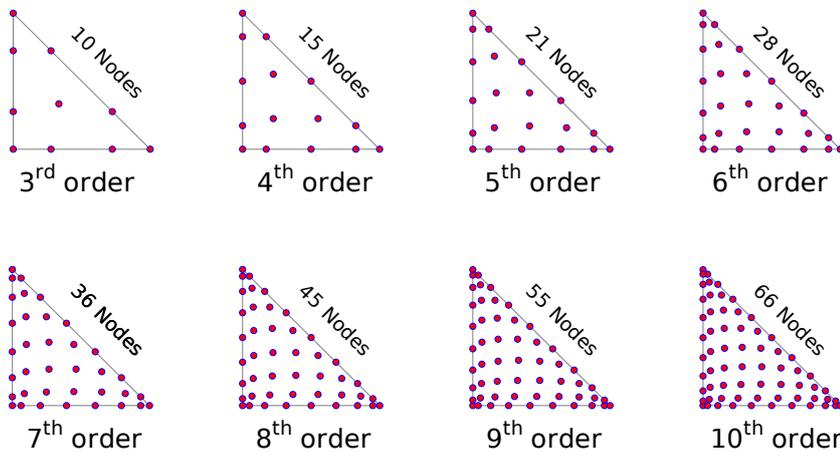
Figure 5.7: *Nodal positions for several expansion orders, obtained via the "warp & blend"-method as proposed in Ref. [47]*
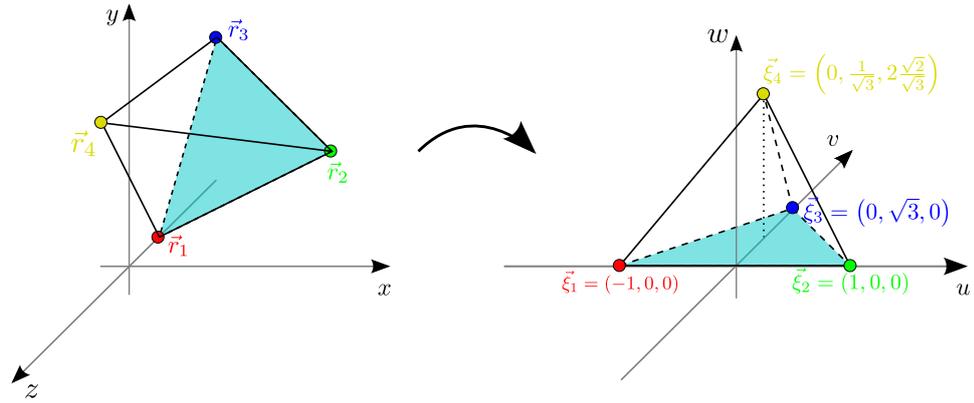
Figure 5.8: *Every tetrahedron is mapped onto the reference tetrahedron with vertices $\vec{\xi}_1$, $\vec{\xi}_2$, $\vec{\xi}_3$, and $\vec{\xi}_4$.*

The corresponding Jacobi matrix can then be obtained from the vertices of the reference tetrahedron $\vec{\xi}_i$ (cf. Fig. 5.8) as

$$
\mathcal{J}^{(3d)} = \begin{pmatrix} x_2 - x_0 & x_3 - x_0 & x_4 - x_0 \\ y_2 - y_0 & y_3 - y_0 & y_4 - y_0 \\ z_2 - z_0 & z_3 - z_0 & z_4 - z_0 \end{pmatrix} \begin{pmatrix} \xi_{2,x} & \xi_{3,x} & \xi_{4,x} \\ \xi_{2,y} & \xi_{3,y} & \xi_{4,y} \\ \xi_{2,z} & \xi_{3,z} & \xi_{4,z} \end{pmatrix}^{-1}
$$

$$
= \begin{pmatrix} x_2 - x_0 & x_3 - x_0 & x_4 - x_0 \\ y_2 - y_0 & y_3 - y_0 & y_4 - y_0 \\ z_2 - z_0 & z_3 - z_0 & z_4 - z_0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{3}} & -\frac{1}{2\sqrt{6}} \\ 0 & 0 & \frac{\sqrt{3}}{2\sqrt{2}} \end{pmatrix}.
$$

In the three-dimensional case, the expansion order N and the number of coefficients N_p is related via

$$
N_p = \frac{(N+1)(N+2)(N+3)}{6}.
$$

Again, we require an orthogonal basis on the reference tetrahedron. For the equilateral tetrahedron pictured in Fig. 5.8, a suitable choice is given by

$$
\Psi_m = \frac{1}{\sqrt{N}} P_i^{(0,0)}(a) P_j^{(2i+1,0)}(b)(1-b)^i P_k^{(2i+2j+2,0)}(c)(1-c)^{i+j}
$$

with

$$
a = \frac{\sqrt{6}u}{\sqrt{6} + \sqrt{2}v + w}, \quad b = \frac{2}{\sqrt{3}} \frac{\sqrt{8}v - w}{\sqrt{8} - \sqrt{3}w} - 1, \quad c = \frac{\sqrt{3}}{\sqrt{2}} w - 1
$$

and the normalization factor

$$
N = \frac{\sqrt{2} \left(4i + 2j + 3\right)^2}{16 \left(1 + \frac{1}{2}\right) \left(i + j + 1\right) \left(i + j + k + \frac{3}{2}\right)}.
$$

Again, $P_i^{(\alpha,\beta)}$ are the Jacobi polynomials and as the special case, $P_i^{(0,0)}$ reduces to the Legendre polynomials.

As in the two-dimensional case, we obtain the missing set of nodes $\vec{r}_i$ via the "warp & blend"-method presented in Ref. [47]. A few exemplary sets of nodes are pictured in Fig. 5.9.
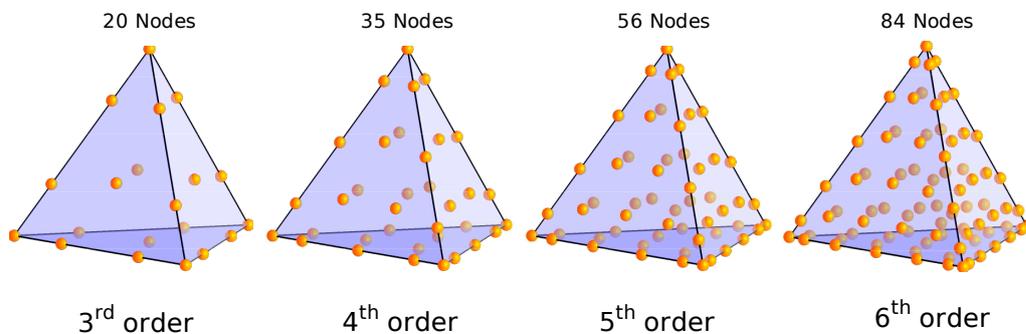
Figure 5.9: *Three-dimensional nodes in the reference tetrahedron as generated by the "warp & blend"-method.*

5.2.5 Putting it all together

With the transformations, the nodal points $\vec{r}_i$ and basis functions Ψ_i at hand, we can continue to simplify the semi-discrete system in Eq. (5.24). In order to do so, we recall the generalized Vandermonde matrix

$$\mathcal{V}_{ij} = \Psi_j(\vec{r}_i).$$

As in the one-dimensional case, this matrix allows us to express the mass-matrix from Eq. (5.25b) as

$$\mathcal{M}^k = J^k \left(\mathcal{V}\mathcal{V}^T\right)^{-1} = J^k \mathcal{M}. \tag{5.28}$$

Similarly, we can rewrite the stiffness-matrices from Eq. (5.25a) as

$$\mathcal{S}_m^k = \left(\mathcal{V}\mathcal{V}^T\right)^{-1} \mathcal{D}_m^k, \quad m = x, y, z \tag{5.29}$$

where $\mathcal{D}_m^k$ denotes the differentiation matrices with respect to the Cartesian coordinates

$$\left(\mathcal{D}_m\right)_{ij} = \sum_k \partial_m \Psi_k(\vec{r}_i)\mathcal{V}_{kj}^{-1}.$$

The main difference to the one-dimensional case results from the matrix $\mathcal{F}$. In order to make this matrix cell-independent, we need to introduce the surface Jacobian J_s^k. In two dimensions, there will be three different values J_s^k per triangle, one corresponding

to each edge. Equivalently, for three-dimensional systems there will be one J_s^k for each face of the tetrahedron. This definition allows to write

$$\mathcal{F}^k = \sum_s J_s^k \mathcal{F}_s, \qquad (5.30)$$

where face matrices $\mathcal{F}_s$ are now given by

$$\left(\mathcal{F}_s\right)_{ij} = \begin{cases} \int_s L_i(\vec{\xi}) L_j(\vec{\xi}) d\vec{\xi}, & \text{where } j \in \{j | \vec{\xi}_j \text{ lies on surface } s\} \\ 0, & \text{otherwise} \end{cases}.$$

As for the mass matrix, one can express the integrals in terms of generalized Vandermonde matrices. However, those matrices must be generated from a basis which is orthogonal on the surface [40].

Inserting the face matrix, together with the matrices from Eq. (5.28) and Eq. (5.29) into Eqs. (5.24) results in the final, semi-discrete scheme

---

Semi-Discrete Discontinuous Galerkin Formulation

$$\partial_t \tilde{\vec{E}}^k = \frac{1}{J^k \epsilon^k} \left[\vec{\mathcal{D}} \times \tilde{\vec{H}}^k + \mathcal{M}^{-1} \sum_s J_s^k \mathcal{F}_s \left(\hat{n} \times \frac{Z^+ \Delta \tilde{\vec{H}}^k - \hat{n} \times \Delta \tilde{\vec{E}}}{Z^- + Z^+} \right) \right], \qquad (5.31a)$$

$$\partial_t \tilde{\vec{H}}^k = \frac{1}{J^k \mu^k} \left[-\vec{\mathcal{D}} \times \tilde{\vec{E}} + \mathcal{M}^{-1} \sum_s J_s^k \mathcal{F}_s \left(\hat{n} \times \frac{-Y^+ \Delta \tilde{\vec{E}} - \hat{n} \times \Delta \tilde{\vec{H}}}{Y^- + Y^+} \right) \right]. \qquad (5.31b)$$

5.3 Convergence and Time-Stepping

Up until now, we have developed the DG scheme mostly based on intuition and mathematical experience. While many steps seem logical, it is not obvious that Eqs. (5.31) present a convergent and stable numerical scheme. Especially the introduction of the numerical flux and its connection to the Riemann problem still lacks a rigorous justification. Furthermore, the semi-discrete scheme in Eqs. (5.31) is only half the way towards a general time-domain solver. It still misses the time-stepping to actually evolve the fields in time.

5.3.1 Convergence

As noted in Sec. 4.2, a rigorous proof of convergence is a condition sine qua non for any numerical method. For the Discontinuous Galerkin method, such a proof can be obtained by making use of the well-known Lax-Richtmyer equivalence theorem [31]. This procedure is very general and rather convenient since it allows to treat the consistency of the spatial discretization independently of the stability of the scheme. Unfortunately, the resulting error estimates are suboptimal [40]. In order to obtain

sharp error bounds, one has to tackle the full problem directly, which was done for Maxwell's equations with the upwind flux in Ref. [19]. In this case, we obtain the error bound

$$\left| \mathbf{q}^{\text{exact}} - \mathbf{q}^{\text{DG}} \right| \leq h^{N+1} \left(c_0 + c_1 T \right), \tag{5.32}$$

where $\mathbf{q}^{\text{exact}}$ denotes an exact solution of Maxwell's equations, while $\mathbf{q}^{\text{DG}}$ represents the numeric approximation. Furthermore, h describes the size of the largest element in the mesh and T is the total time of simulation.

From this result, we can deduce two important points:

- The deviation of our calculations from the exact result does not grow faster than linearly with time. As we will demonstrate later, this is a sharp result.

- The slope of the growth is determined by the mesh size h and the polynomial order N. Here, the convergence is algebraic in h and exponential in N. This feature is also known as hp-convergence.

We will return to the point of accuracy in Sec. 5.5, where the performance of the scheme will be analyzed in more detail.

5.3.2 Time-Stepping

The final missing component in deriving a time-domain solver from the semi-discrete system (5.31) is the integration in time. The simplest possible solution would be to employ a finite-difference discretization as we did for the FDTD method. Unfortunately, this method shows a rather poor performance and would limit the efficiency of our solver dramatically. Instead, we adopt a more sophisticated integration scheme known as the low-storage Runge-Kutta (LSRK) method.

In contrast to classical Runge-Kutta schemes, which are known for their versatility and excellent performance, the LSRK methods are tuned to require less memory. This feature becomes important, when we try to simulate a very large system where the number of unknowns reaches the order of $M \approx 10^7$ or more. But as we will demonstrate later on, the LSRK methods are also advantageous for smaller systems.

If we rewrite Eqs. (5.31) as

$$\frac{\partial}{\partial t} \mathbf{u}(t) = f(\mathbf{u}, t), \tag{5.33}$$

then we can integrate it in time by using the $2M$-LSRK due to Williamson [49] as

$$\mathbf{y}_0 = \mathbf{y}(t)$$
$$\left. \begin{array}{l} \mathbf{k}_i = \tilde{a}_i \mathbf{k}_{i-1} + \Delta t f(\mathbf{y}_{i-1}, t + \tilde{c}_i \Delta t) \\ \mathbf{y}_i = \mathbf{y}_{i-1} + \tilde{b}_i \mathbf{k}_i \end{array} \right\} \quad \forall i = 1 \ldots s$$
$$\mathbf{y}(t + \Delta t) = \mathbf{y}_s.$$

Here, s denotes the number of stages while $\tilde{a}_i$, $\tilde{b}_i$ and $\tilde{c}_i$ are the coefficients which define a particular scheme. In Ref. [50], Carpenter and Kennedy propose a set of coefficients

i	$\tilde{a}_i$	$\tilde{b}_i$	$\tilde{c}_i$
1	0	$\frac{1432997174477}{9575080441755}$	0
2	$-\frac{567301805773}{1357537059087}$	$\frac{5161836677717}{13612068292357}$	$\frac{1432997174477}{9575080441755}$
3	$-\frac{2404267990393}{2016746695238}$	$\frac{1720146321549}{2090206949498}$	$\frac{2526269341429}{6820363962896}$
4	$-\frac{3550918686646}{2091501179385}$	$\frac{3134564353537}{4481467310338}$	$\frac{2006345519317}{3224310063776}$
5	$-\frac{1275806237668}{842570457699}$	$\frac{2277821191437}{14882151754819}$	$\frac{2802321613138}{2924317926251}$

Table 5.3: *Coefficients for a low-storage Runge-Kutta scheme according to Ref. [50]. The scheme requires $s = 5$ stages and is fourth-order accurate in time.*

for a suitable integration scheme of fourth order. Unfortunately, in contrast to the classical Runge-Kutta methods, it is not possible to create such a scheme with only four stages. Instead one requires at least $s = 5$. The proper coefficients are given in Tab. 5.3.

As an explicit scheme, the LSRK scheme is only conditionally stable. Similarly to the case of FDTD, we need to fulfill a CFS-criterion which limits the allowed timestep. However, in contrast to FDTD, we can not express this criterion explicitly. Instead, we can find the appropriate condition by considering the ordinary differential equation

$$\frac{\partial}{\partial t} y = \lambda y. \tag{5.34}$$

Inserting this equation into the Runge-Kutta yields

$$y(t + \Delta t) = \Delta t \underbrace{\left(1 + \alpha_1 \lambda + \alpha_2 \lambda^2 + \cdots + \alpha_s \lambda^s \right)}_{=:G(\lambda)} y, \tag{5.35}$$

where α_n are numerical coefficients depending on the parameters $\tilde{a}_i$ and $\tilde{b}_i$ of the particular scheme. Obviously, for our solution not to be growing we must fulfill the condition

$$\Delta t \, G(\lambda) \leq 1. \tag{5.36}$$

A contour of the function $G(\lambda) = 1$ is plotted in Fig. 5.10 for a classical fourth-order Runge-Kutta scheme and for our LSRK integrator. Clearly, the LSRK has a significantly larger region of stability. Besides the lower memory requirements, this is the reason for employing a LSRK method instead of a classical Runge-Kutta integrator. Returning to our full problem, λ plays the role of to the eigenvalues of the discrete operator on the r.h.s. of Eqs. (5.31). Thus, we are required to pick Δt sufficiently small to ensure that the largest eigenvalue still fits into the region of stability.

Choosing Δt in Practice

Unfortunately, we can not afford to calculate the spectrum of each operator before we conduct a time-domain calculation. Therefore, we need a more accessible way to obtain
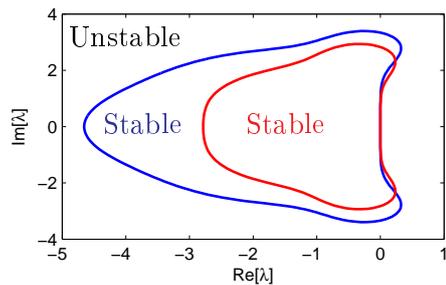
Figure 5.10: *Stability region of the classical Runge-Kutta (red) and a LSRK scheme (blue).*

a sufficiently small Δt without wasting computational resources by picking it too small. Empirically, we know that the largest eigenvalue is connected to the shortest distance between to gridpoints in our system. This scale can be separated into two parts. First, we have the nodes on an individual elements. As a rather primitive measure for the minimal distance between them, we consider the LGL-nodes along the edges. There, we easily calculate the distance $\Delta u_N = u_2 - u_1$ of the first two nodes on the unit interval. This length will depend on the expansion order N as $\Delta u_N \propto N^{-2}$. Further, we need a measure for the size of the smallest element. Following the suggestions of Hesthaven and Warburton [40], we pick the incircle or insphere of the triangles or tetrahedra, respectively. This length is denoted by r^k for each element k. Then, we determine the timestep as

$$\Delta t = s \, \Delta u_N \, \min_{\Omega} \left(r^k \right), \tag{5.37}$$

where s is a number of order 1, which we determine from a set of numerical experiments. We will return to this point in Sec. 5.5.

5.4 Sources, Dispersive Media and PMLs

Before we proceed to test the DG method numerically, we briefly comment on some of the relevant extensions.

Sources

A first important point is the excitation of the system through sources. Besides via initial conditions, there are two conceptually different ways to inject radiation into the system. The more obvious path is to add current density terms to the r.h.s. of Eqs. (5.31). An alternative is given by the so-called Total-Field/Scattered-Field approach as discussed in Sec. 4.3.2. Typically, the current density is more suitable to introduce localized sources while the TF/SF method can be employed to inject plane waves. Within the DG framework, one can easily implement the TF/SF approach by modifying the field differences in the numerical fluxes. The addition of current terms

also presents no fundamental problem, but the spatial profile is obviously expanded into Lagrange polynomials. For the very important case of a delta-like point source, this expansion becomes rather intricate and a highly refined mesh around the source is required to model it accurately. In turn, this dramatically reduces the computational performance due to the stability constraints on the timestep. Therefore, we implement point sources by exploiting the TF/SF method. For the injection, a small contour around the desired source location is embedded into the mesh and the outer area is defined as the TF region. The required fields on the contour can be obtained by means of Green's functions as presented in Sec. 2.5.

Dispersive Media

A second extension is the implementation of dispersive materials. As discussed in Sec. 3.1.3, we can include certain models of dispersion via auxiliary differential equations (ADEs). Since the resulting ADEs for typical dispersion models do not contain spatial derivatives, we can easily incorporate them into the time integration. It was shown in Ref. [51], that additional ADEs without spatial derivatives do not influence the choice of the numerical flux.

Perfectly Matched Layer

Finally, we return to the crucial point of PMLs. We have seen in Sec. 3.2, that one can implement PMLs in multiple formulations. The difference between the two presented versions, namely UPMLs and CFS-PMLs, lies in the number and form of the resulting ADEs. As demonstrated, a UPML formulation can be implemented by a single ADE per field component, which does not contain additional spatial derivatives. Unfortunately, this is not true for the more general CFS-PMLs. Here, we require two additional fields and the equations contain additional spatial derivatives. As a consequence, we abstain from implementing CFS-PMLs and employ standard UPMLs instead. In Sec. 5.5.3, we will demonstrate that one can still achieve excellent absorption properties by optimizing the free parameters.

5.5 Verification and Comparison to FDTD

Similarly to the verification of the FDTD method in Sec. 4.4, we employ a number of basic systems to test the performance and stability of our implementation. To allow for direct comparison, we start with the same tests as for the FDTD case.

5.5.1 Empty Metallic Cavities

In contrast to the FDTD method, the performance DG scheme is affected by a number of external parameters. Particularly the shape and quality of the mesh can significantly influence the results. For this section, we employ regular meshes with PMC boundary conditions as depicted in Fig. 5.11. As a first test, we return to the question
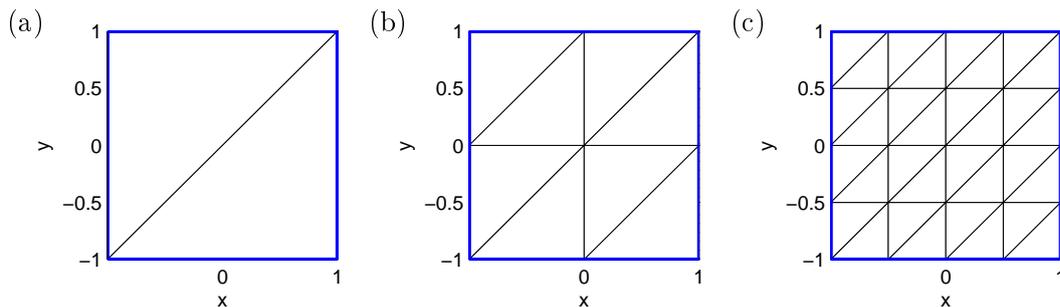
Figure 5.11: *Regular test meshes for an empty, metallic cavity. (a), (b) and (c) depict meshes for $h = 1$, $h = \frac{1}{2}$ and $h = \frac{1}{4}$, respectively. The blue frame indicates PMC boundary conditions.*

of the proper timestep Δt. According to the discussion in Sec. 5.3.2, we still need to empirically determine an appropriate value for the factor s in Eq. (5.37). To do so, we simulate the cavity for different meshes, different orders N and different values for s. The resulting maximal error after $T = 50\,T_0$ is plotted as a function of s in Fig. 5.12. From these results, one can directly extract the point of instability. Fur-
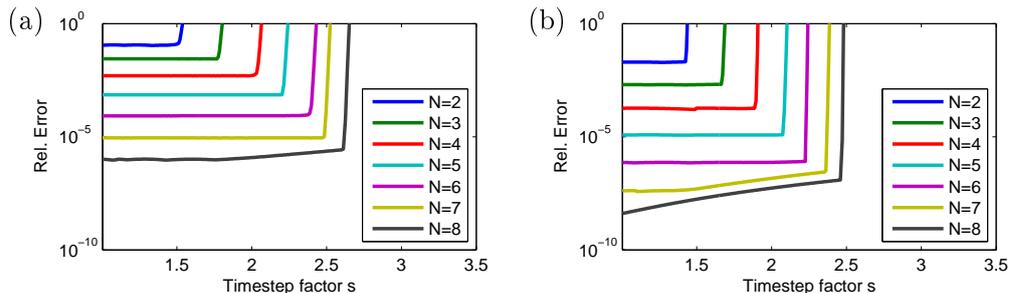
Figure 5.12: *Error for empty cavity calculations as a function of the timestep. The data in (a) was obtained with 4 triangles per wavelength, while (b) contains data for 8 triangles per wavelength.*

ther, we find that once the scheme is stable, most of the results do not change with the timestep. Thus, we conclude that the error is mostly dominated by the spatial discretization. Only for relative errors below 10^{-5} one can observe a further reduction of the error with decreasing timesteps. As a consequence, we will generally try to make the timestep as large as possible without crossing the instability limit. Only for highly accurate calculations, one has to slightly reduce s. From the data in Fig. 5.12, we find that the maximal value $s_{\max}$ still depends on both, the spatial order and the size of the mesh. In Fig. 5.13 we plot the maximal values of $s_{\max}$ for different orders and meshes. We observe how the maximally allowed timestep factor increases with the order and decreases with refinement of the mesh. The variations, however, are not too large,
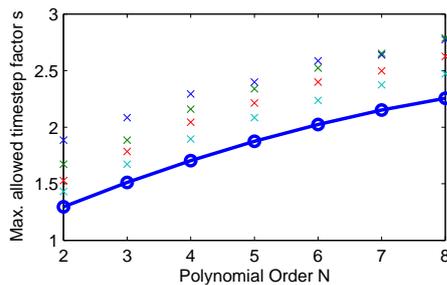
Figure 5.13: *The crosses indicate maximally allowed timestep factors s_{max} for different meshes and orders. The solid line corresponds to values for s_{max} as given by Eq. (5.38).*
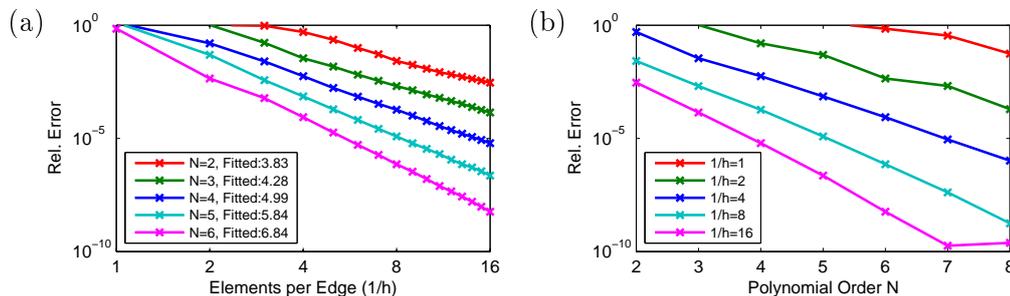
Figure 5.14: *(a) Relative error as a function of mesh refinement. Straight lines in the semi-logarithmic plot indicate algebraic convergence. The fitted slopes are given in the legend. (b) Relative error as a function of increasing order of the polynomial expansion. The straight lines in this double logarithmic plot demonstrate exponential convergence.*

indicating that the scaling discussed in Sec. 5.3.2 works reasonably well. In order to increase performance, we fit a quadratic polynomial to the maximally allowed timestep for each order of the smallest mesh. To be on the save side, we then subtract a certain margin of error and obtain the expression

$$s_{max}(N) = 0.8 + 0.27N - 0.011N^2. \tag{5.38}$$

The resulting values are indicated in Fig. 5.13 as blue circles. Those values will be used for all realistic calculations, unless stated otherwise.

However, before we proceed to realistic setups, we should also analyze the errors as a function of the mesh and the spatial order. To completely eliminate the influence of the time integration, we will work with a timestep factor of $s = 0.1 s_{max}$. In Fig. 5.14, we display the error as a function of the mesh size (Fig. 5.14(a)) and of the polynomial order (Fig. 5.14(b)). These results nicely confirm the theoretical analysis of the error bounds presented in Sec. 5.3.1. In particular, from the straight lines in the semi-logarithmic plot 5.14(a) we confirm the algebraic h-convergence. Conducting a linear
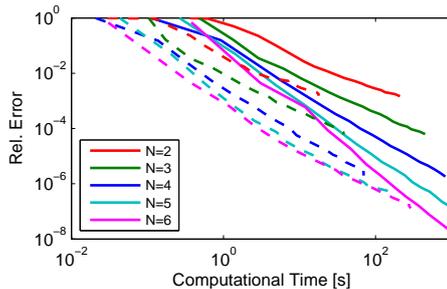
Figure 5.15: *The relative error as a function of the computational time. The dashed lines correspond to calculations with a timestep factor of $s = s_{\max}$, while the data for the solid lines was obtained with $s = 0.1s_{\max}$.*

fit, we find the slope roughly equivalent to $N+1$, as predicted by Eq. 5.32. Furthermore, we observe exponential p-convergence from the double logarithmic plot in Fig. 5.14(b).

The qualitatively different convergence behavior naturally leads to the question of which refinement should be preferred. To this end, we plot the relative error as a function of the computational time spent to obtain the result in Fig. 5.15. From these results, we observe that for this problem the p-refinement does indeed result in faster calculations for a given error. But it should be noted, that this can not directly be transfered to calculations of realistic systems. Particularly if curved interfaces are involved, the error can be strongly dominated by the approximation due to straight-sided elements. In this case, only h-refinement will lead to any improvement. Similar considerations are also valid, if we have to deal with field singularities.

5.5.2 Half-Filled Cavities

As a next step we demonstrate the superior convergence properties of the DG methods, when interfaces are involved. In Sec. 4.4.2 we demonstrated that a fourth-order FDTD scheme is reduced to second order of accuracy if material interfaces are present. Now we demonstrate that the DG method does not exhibit the same restriction. The physical setup is identical to the one in Sec. 4.4.2 and the corresponding meshes are shown in Fig. 5.16. As in the previous section, we employ time-step of $s = 0.1s_{\max}$ to avoid error contributions from the time integration. The important feature is that both, h- and p-convergence remain entirely unaffected by the interface (cf. Fig. 5.17). This is in stark contrast to the FDTD results and highlights the versatility of the DG method.

Finally, we conduct a direct comparison with the FDTD method. To this end, we plot the DG results obtained with the maximal timestep factor as a function of CPU time in Fig. 5.18. In the same plot, we also depict the computational time required for the corresponding FDTD calculations. From this plot, we observe that the DG method is clearly superior for the given problem. If we require a relative error below 1%, then DG delivers the solution about two orders of magnitude faster than FDTD. However, this advantage largely stems from the higher-order nature of our method. As
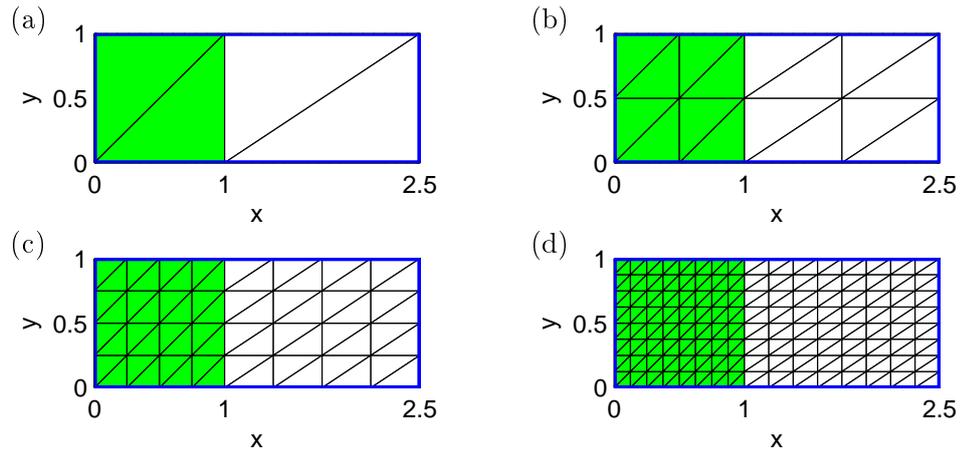
Figure 5.16: *Meshes with different levels of refinement for the semi-filled cavity* $(h = 1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8})$. *The green shaded triangles mark the region filled with* $\epsilon = 2$.
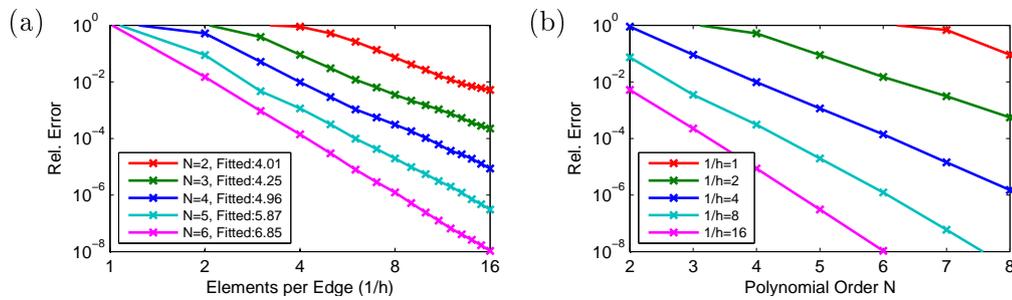
Figure 5.17: *(a) Relative error as a function of mesh refinement. Straight lines in the semi-logarithmic plot indicate algebraic convergence. The fitted slopes are given in the legend. (b) Relative error as a function of increasing order of the polynomial expansion. The straight lines in this double logarithmic plot demonstrate exponential convergence.*
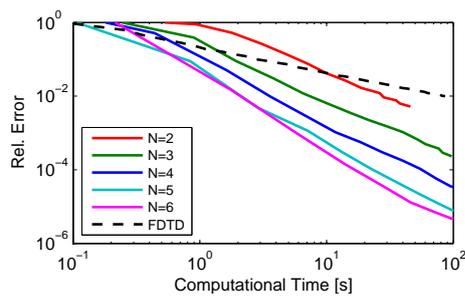
Figure 5.18: *The relative error as a function of the computational time. The solid lines correspond to DG calculations with a timestep factor of* $s = s_{\max}$, *while the dashed line represents results from an FDTD calculation.*

(a)   (b)
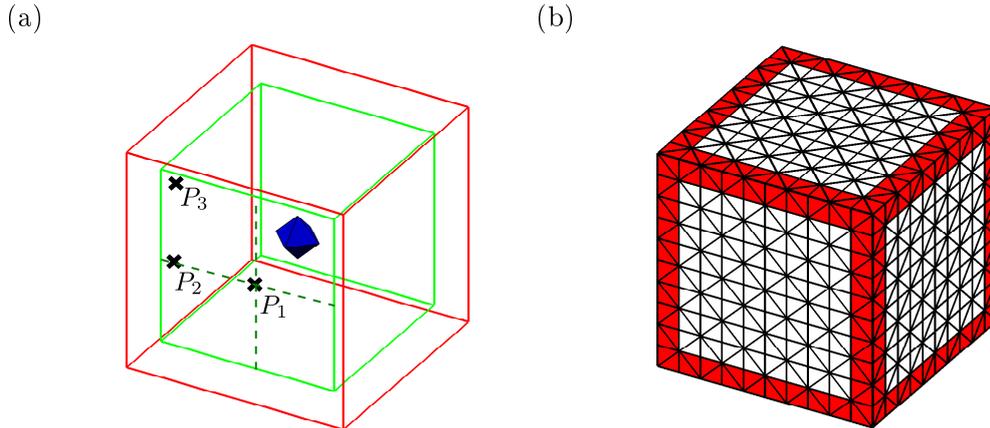
Figure 5.19: *System used for the validation of PMLs. (a) shows the interface between the computational region and the PMLs (green frame) as well as the outer boundaries (red frame). The blue surface in the center marks the contour used to inject the point source, while the crosses indicate the recording points. (b) depicts the tetrahedral mesh. Edges and Corners are shaded red to indicate the extent of the PML region (one cell).*

discussed previously, for more complicated structures containing curved interfaces, we might not be able to fully exploit the *p*-convergence. For those cases, the advantage of the DG method is diminished but usually still existent.

5.5.3 Optimization of the PML Parameters

Finally, we revisit the problem of absorbing boundary conditions. As alluded to above, the PMLs still contain a free parameter σ which needs to be determined numerically for best performance. In the case of FDTD, we have already demonstrated that σ must be made position dependent, to optimize absorption (see Sec. 4.4.3). There, we chose a polynomial grading

$$\sigma(x) = (x/d)^m \frac{(m+1)}{2d} R,$$

where d is the thickness of the PML-layer while m and R are free parameters. Best results for FDTD are obtained for values of $m \approx 3$ and $R \approx 10$. To facilitate a comparison of PMLs in the DG method with those of FDTD, we adopt the same polynomial grading with free parameters m and R. In order to scan the entire parameter space, we use a small three-dimensional test system as sketched in Fig. 5.19. A point-source is placed in the center of the system and is injected via a TF/SF contour as discussed in Sec. 5.4. The pulse shape, the definition of the error and the recording positions are all chosen identically to the FDTD case (cf. Sec. 4.4.3). As before, we compare with numerical results obtained on a much larger system.

In Fig. 5.20, we depict the results of our parameter studies for the system sketched in Fig. 5.19. Most importantly, we note that for $N = 3$ and only one layer of PMLs, we
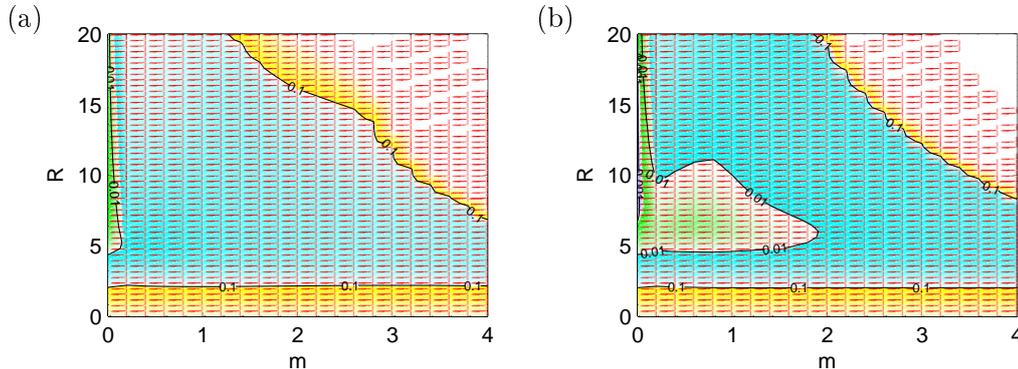
Figure 5.20: *Dependence of the relative error on the parameters m and R for spatial order (a) N=3 and (b) N=4. In both cases, the error of the field at recording point P_1 is plotted.*

already achieve a relative error below 1%. Moreover, this minimum does not coincide with the one expected from FDTD. Instead, the best performance can be observed for $m = 0$ which renders the grading of σ obsolete. This result is not so surprising and fully consistent with our results for a two-dimensional test case [52].

Instead of wondering why the grading does not help, we have to ask why it was needed in the first place. As we have stated previously, there is no analytic reason for making σ spatially dependent, since the PMLs should be reflectionless for all values of σ. Instead, we had to introduce the grading because the FDTD method does not deal well with discontinuities in the material parameters. Therefore, an adiabatic ramping can help to mitigate numerical problems with the derivatives across the PML interface. Since the DG method by construction has no problems with discontinuous fields, there is no longer a reason for the grading.

In a second step, we now have to search for the optimal value of R, which defines the damping inside the PMLs. Analytically, this value should be as large as possible to prevent radiation making a round-trip through the PML and returning to the computational domain. Our results of a detailed scan are presented in Fig. 5.21. As expected theoretically, the error diminishes exponentially with increasing R until a certain point. At this point, the damping gets so strong, that we fail to properly approximate the fields inside the PML with a polynomial of the given order. Then, the error increases again, because the approximation of the fields gets worse. This argument is supported by the observation, that the error reduces further if the order is increased. Thus, the optimal value for R depends on the polynomial order used. For our system, the optimal value lies around $R \approx 10$. This is the value that will be employed in all further calculations.
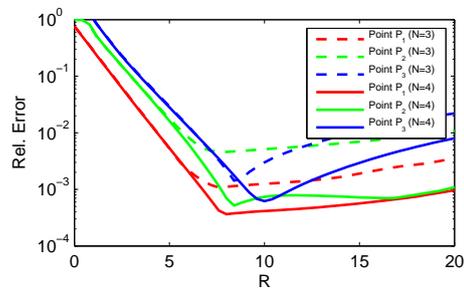
Figure 5.21: *Performance of the PMLs for $m = 0$ as a function of the param-eter R. The colors denote the three different recording points, while the linestyles correspond to N=3 (solid) and N=4 (dashed).*

6 Chapter 6

Advanced Time Integration

So far, we were mostly concerned with the spatial discretization of Maxwell's equations while the time integration was only touched on. For both the standard FDTD formulation and for most discontinuous Galerkin calculations, this was justified because the errors are strongly dominated by the spatial discretization. Only for highly resolved systems we could observe a degradation of the accuracy due to insufficient time integration. But besides the point of accuracy the time integration obviously has a strong influence on the performance of any time-domain method. All schemes discussed so far are fully explicit, which means that they allow a direct computation of the following timestep from the previous one. Unfortunately, those methods all suffer from performance restrictions due to the Courant-Friedrichs-Lewy (CFL) criterion (cf. Sec. 4.2.1 and Sec. 5.3.2). To overcome this limitation, one could employ implicit integrators, but they require the numerical solution of a large system of linear equations for each timestep. This additional computational effort often negates the performance gains due to the larger timesteps.

In this chapter, we present an alternative class of time integration methods, which can not be attributed to either explicit or implicit schemes. Those methods are called exponential integrators and we will demonstrate that they can exhibit unconditional stability without the computational effort usually connected with an implicit integrator. As we will see, these methods also provide a systematic way to increase the accuracy of the time integration.

6.1 Linear Matrix Exponential Integrators

To introduce the idea behind matrix exponential integrators, we start from Maxwell's curl equations (2.10) with linear, non-dispersive media and without explicit source terms. Combining the electric and magnetic fields into a supervector $\mathbf{q} = \left(\vec{E}, \vec{H} \right)^{\mathrm{T}}$ allows us to rewrite the curl equations as

$$\frac{\partial}{\partial t}\mathbf{q} = \hat{\mathcal{H}}\mathbf{q}, \tag{6.1}$$

where the operator $\hat{\mathcal{H}}$ is given by

$$\mathcal{H} = \begin{pmatrix} 0 & \epsilon(\vec{r})^{-1}\nabla\times \\ -\mu(\vec{r})^{-1}\nabla\times & 0 \end{pmatrix}. \tag{6.2}$$

By formal integration, a solution to Eq. (6.1) is readily found as

$$\mathbf{q}(t_0 + \Delta t) = \exp(\hat{\mathcal{H}}\,\Delta t)\mathbf{q}(t_0). \tag{6.3}$$

As discussed in the previous chapters, spatial discretization of the fields will convert the operator $\hat{\mathcal{H}}$ into a finite matrix $\mathcal{H}$. Thus, the problem of time integration is transformed into the task to evaluate the exponential function of a matrix. While there is a large variety of methods to calculate such matrix exponentials [53], only few of them are suitable for our task.

The particular problem becomes apparent when we consider an exemplary three-dimensional system, discretized with finite differences. Assuming a discretization of 100 cells along each axis results in a system of approximately $N_{\text{tot}} = 6\times(100\times100\times100) = 6\times10^{6}$ unknowns. While the storage of $N_{\text{tot}} \approx 10^{7}$ variables presents no challenge for modern computers, the corresponding matrix $\mathcal{H}$ would have $N_{\text{tot}}^{2} \approx 10^{14}$ entries. This amount of data can neither be stored nor processed on even the most advanced machines currently available.

The key to the solution of this problem are two points. First, for most discretizations the matrix $\mathcal{H}$ will be sparse, i.e., most entries are zero. This dramatically reduces the required storage and it enables us to quickly evaluate matrix-vector products $\mathcal{H}\mathbf{q}$. Unfortunately, the sparsity of $\mathcal{H}$ does not imply a sparse matrix $\exp\big(\mathcal{H}\,\Delta t\big)$. The second point is therefore, to realize that the explicit result $\exp\big(\mathcal{H}\,\Delta t\big)$ is not required to do the time integration. Instead, we only need the action of this matrix onto a given initial state $\mathbf{q}(t_0)$. Thus, we are looking for a method which evaluates the action of the matrix $\exp\big(\mathcal{H}\,\Delta t\big)$ on a vector $\mathbf{q}(t_0)$ without the requirement to store either $\mathcal{H}$ or $\exp\big(\mathcal{H}\,\Delta t\big)$ explicitly.

The most obvious approach would be to exploit the Taylor expansion

$$\exp(\mathcal{H}\,\Delta t)\mathbf{q}(t_0) = \sum_{n=0}^{\infty} \frac{(\mathcal{H}\,\Delta t)^n}{n!}\mathbf{q}(t_0),$$

and truncate the series at a certain order. Unfortunately, the numerical convergence of the Taylor series is known to be very poor [53]. Recently, two alternative schemes to evaluate a matrix exponential with respect to Maxwell's equations were published by De Raedt and coworkers [54], but both of them rely on the skew-symmetry of $\hat{\mathcal{H}}$. This restriction severely limits the flexibility of those methods and renders them unsuitable for extensions to nonlinear solvers [55]. In the following sections, we will present a method based on the so-called Krylov subspace approach which is free of any requirements on the structure of the matrix.

$\mathbf{q}_1 \leftarrow \mathbf{q}(t_0)/\left|\mathbf{q}(t_0)\right|$
`for` $j = 1 \ldots m$
$\mathbf{w} \leftarrow \mathcal{H}\mathbf{q}_i$
`for` $i = 1 \ldots j$
$h_{i,j} = \mathbf{w} \cdot \mathbf{q}_i$
$\mathbf{w} \leftarrow \mathbf{w} - h_{i,j}\mathbf{q}_i$
$h_{i+1,j} = \left|\mathbf{w}\right|$
$\mathbf{q}_{j+1} \leftarrow h_{i+1,j}$
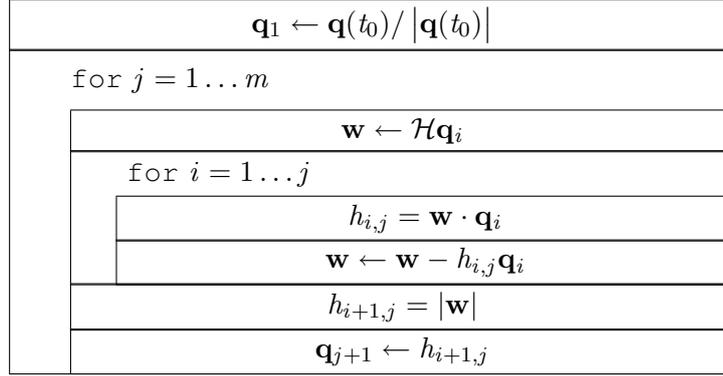
Figure 6.1: *Schematic representation of the Arnoldi algorithm as a Nassi-Shneiderman diagram.*

6.1.1 Krylov Subspace Techniques

The basic idea behind any Krylov-subspace method is to project a large, sparse matrix onto a much smaller, suitably chosen subspace. In this subspace, the actual calculation of the matrix exponential (or any other matrix operation) is performed. Projecting back the result then yields an approximation of the operation on the full matrix [56]. The m-dimensional Krylov subspace K_m is defined by

$$K_m = \text{span}\{\mathbf{q}(t_0), \mathcal{H}\mathbf{q}(t_0), \mathcal{H}^2\mathbf{q}(t_0), \ldots, \mathcal{H}^{m-1}\mathbf{q}(t_0)\}, \tag{6.4}$$

where the subspace dimension m is typically chosen to be of the order of ten and, therefore, many orders of magnitude smaller than the number of unknowns N_{tot}. In a next step, we need to execute the actual projection of $\mathcal{H}$ onto the subspace K_m. For general matrices, this can be achieved by the Arnoldi algorithm [57] as sketched in Fig. 6.1. It should be noted, that the inner loop of said algorithm is nothing but a modified Gram-Schmidt process, which leads to an orthonormal basis

$$\mathcal{V}_m = \big[\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_m\big]$$

of K_m. Furthermore, the algorithm generates a matrix H_m as

$$\big(\mathcal{H}_m\big)_{ij} = h_{i,j}, \quad i, j \leq m; i \leq j + 1,$$

which is of upper Hessenberg form. By generation, the matrix $\mathcal{H}_m$ fulfills the relation

$$\mathcal{H}\mathcal{V}_m = \mathcal{V}_m\mathcal{H}_m + h_{m+1,m}\mathbf{q}_{m+1}\mathbf{e}_m^{\text{T}}. \tag{6.5}$$

Here, $\mathbf{e}_i$ denotes the i-th unit vector belonging to the space $\mathbb{R}^m$. Upon multiplication of Eq. (6.5) with $\mathcal{V}_m^{\text{T}}$, we obtain

$$\mathcal{H}_m = \mathcal{V}_m^{\text{T}}\mathcal{H}\mathcal{V}_m.$$

Therefore, $\mathcal{H}_m$ represents the projection of $\mathcal{H}$ onto the subspace K_m, with respect to the basis $\mathcal{V}_m$ [56, 58].

In a next step, we note that the best approximation $\mathbf{q}_{\text{opt}}(t_0 + \Delta t)$ to $\exp\left(\mathcal{H}\Delta t\right)\mathbf{q}(t_0)$ in the subspace K_m is given by the projection [58]

$$\mathbf{q}_{\text{opt}}(t_0 + \Delta t) = \mathcal{V}_m \mathcal{V}_m^{\text{T}} e^{\mathcal{H}\Delta t}\mathbf{q}(t_0).$$

Exploiting the identities $\mathbf{q}_1 = \mathbf{q}(t_0)/\left|\mathbf{q}(t_0)\right|$ and $\mathbf{q}_1 = \mathcal{V}\mathbf{e}_1$, it follows that

$$\mathbf{q}_{\text{opt}}(t_0 + \Delta t) = \left|\mathbf{q}(t_0)\right| \mathcal{V}_m \underbrace{\mathcal{V}_m^{\text{T}} e^{\mathcal{H}\Delta t}\mathcal{V}_m}_{=:\mathcal{L}}\mathbf{e}_1.$$

The matrix $\mathcal{L}$ is a dense $m \times m$ matrix, but unfortunately it is not practically computable, since it still contains the full matrix exponential $e^{\mathcal{H}\Delta t}$. However, a Taylor expansion of $\mathcal{L}$ around $\Delta t = 0$ results in

$$\begin{aligned}
\mathcal{L} &= \mathcal{V}_m^{\text{T}} \sum_{n=0}^{\infty} \frac{\left(\mathcal{H}\Delta t\right)^n}{n!}\mathcal{V}_m \\
&= \mathcal{V}_m^{\text{T}} \sum_{n=0}^{m-1} \frac{\left(\mathcal{H}\Delta t\right)^n}{n!}\mathcal{V}_m + \mathcal{V}_m^{\text{T}} \sum_{n=m}^{\infty} \frac{\left(\mathcal{H}\Delta t\right)^n}{n!}\mathcal{V}_m \\
&= \mathcal{V}_m^{\text{T}} \sum_{n=0}^{m-1} \frac{\left(\mathcal{V}_m \mathcal{V}_m^{T}\mathcal{H}\Delta t\right)^n}{n!}\mathcal{V}_m + \mathcal{V}_m^{\text{T}} \sum_{n=m}^{\infty} \frac{\left(\mathcal{H}\Delta t\right)^n}{n!}\mathcal{V}_m,
\end{aligned}$$

where we have exploited the fact that $\mathcal{V}_m\mathcal{V}_m^{T}\mathcal{H}^n = \mathcal{H}^n$ for all $n \leq m$, by construction. Expanding the first sum as

$$\left(\mathcal{V}_m\mathcal{V}_m^{T}\mathcal{H}\Delta t\right)^n = \left(\mathcal{V}_m\underbrace{\mathcal{V}_m^{T}\mathcal{H}\mathcal{V}_m}_{\mathcal{H}_m}\mathcal{V}_m^{T}\mathcal{H}\cdots\mathcal{V}_m\mathcal{V}_m^{T}\mathcal{H}\right)\Delta t^n = \mathcal{V}_m\left(\mathcal{H}_m\Delta t\right)^n$$

then yields the expression

$$\mathcal{L} = \mathcal{V}_m^{\text{T}} e^{\mathcal{H}\Delta t}\mathcal{V}_m = e^{\mathcal{H}_m\Delta t} + O\left(\Delta t^m\right). \tag{6.6}$$

This allows us to formulate the main approximation of the linear Krylov-subspace scheme as

> **Main Approximation**
> $$e^{\mathcal{H}\Delta t}\mathbf{q}_0 \approx \left|\mathbf{q}_0\right| \mathcal{V}_m e^{\mathcal{H}_m\Delta t}\mathbf{e}_1. \tag{6.7}$$

As we have seen from Eq. (6.6), this approximation is at least accurate up to $O\left(\Delta t^m\right)$ in time. A more accurate error bound is given by [56]

$$\left|e^{\mathcal{H}\Delta t}\mathbf{q}(t_0) - \left|\mathbf{q}(t_0)\right| \mathcal{V}_m e^{\mathcal{H}_m\Delta t}\mathbf{e}_1\right| \leq 2\left|\mathbf{q}(t_0)\right| \frac{\rho^m e^{\rho}}{m!},$$

where $\rho = \left| \mathcal{H} \, \Delta t \right|$, and $|\cdot|$ denotes the standard two-norm of a vector or the spectral norm of a matrix, respectively. Thus, for sufficiently large m, we can expect the $\frac{1}{m!}$-dependence to dominate, leading to a superlinear convergence of the scheme.

Exponential of Small Matrices

After establishing the general procedure of a Krylov-subspace based exponential integrator, we now need to address the problem of calculating $e^{\mathcal{H}_m \Delta t}$, where $\mathcal{H}_m$ is a small upper Hessenberg matrix. As alluded to above, there is a variety of methods for calculating the exponential of a matrix [53]. For our particular problem, the class of rational approximations

$$e^{\mathcal{H}_m \Delta t} \approx \frac{P_i \left(\mathcal{H}_m \Delta t \right)}{Q_j \left(\mathcal{H}_m \Delta t \right)}$$

has proven itself successfully. Here, P_i and Q_j are polynomials of degrees i and j respectively. In particular, we employ a (n,n) Padé approximation, which takes the form [59]

$$e^{\mathcal{H}_m \Delta t} \approx \frac{P_n \left(\mathcal{H}_m \Delta t \right)}{P_n \left(-\mathcal{H}_m \Delta t \right)},$$

where

$$P_n \left(\mathcal{H}_m \Delta t \right) = \sum_{i=0}^{n} c_i \left(\mathcal{H}_m \Delta t \right)^i .$$

The corresponding coefficients c_i are given by the recursion

$$c_i = \begin{cases} 1, & \text{if } i = 0 \\ c_{i-1} \frac{n-i+1}{(2n-i+1)i}, & \text{if } i > 0 \end{cases} .$$

In practice, an order of $n = 6$ was found to be sufficient to achieve converged results.

6.1.2 Implementation of Sources

So far, our discussion was restricted to systems without sources, which clearly is of limited practical use. A generalization of Eq. (6.1) with respect to source terms can be expressed as

$$\frac{\partial}{\partial t} \mathbf{q} = \hat{\mathcal{H}} \mathbf{q} + \mathbf{j}(t), \tag{6.8}$$

where the supervector $\mathbf{j}(t) = \left(\vec{j}_E, \vec{j}_H \right)^{\mathrm{T}}$ now contains the current source terms. Similarly to Eq. (6.1), we can formally integrate Eq. (6.8) and find the solution by variation of constants as

$$\mathbf{q}(t_0 + \Delta t) = e^{\mathcal{H} \, \Delta t} \mathbf{q}(t_0) + \int_{t_0}^{t_0 + \Delta t} e^{\left(t_0 + \Delta t - \tau \right) \mathcal{H}} \mathbf{j}(\tau) d\tau. \tag{6.9}$$

Thus, the solution of a system with sources requires us to solve the source-free problem and, additionally, we have to evaluate a convolution integral containing matrix exponentials.

Direct Integration via Quadrature Rules

A very general procedure of handling the convolution integral in Eq. (6.9) can be found by recasting the problem into the time-stepping procedure [58]

$$\mathbf{q}(t_0 + \Delta t) = \mathbf{q}(t_0) + \int\limits_0^{\Delta t} e^{(\Delta t - \tau)\mathcal{H}} \underbrace{\left[\mathbf{j}(t_0 + \tau) + \mathcal{H}\mathbf{q}(t_0)\right]}_{\boldsymbol{\Phi}_0(\tau)} d\tau.$$

Upon replacing the integral with a suitably chosen quadrature formula, one obtains

$$\mathbf{q}(t_0 + \Delta t) = \mathbf{q}(t_0) + \sum_{i=1}^n w_i\, e^{(\Delta t - \tau_i)\mathcal{H}} \boldsymbol{\Phi}_0(\tau_i). \tag{6.10}$$

Here, τ_i and w_i ($i = 1, \ldots, n$) are, respectively, the nodes and associated weights of the quadrature scheme. Note that for an n-point quadrature we would require n Krylov-subspace calculations. In turn, this would dramatically decrease the performance of the method. Instead, Gallopoulos and Saad [58] proposed to work in only one subspace, which, for instance, may be built from the vector $\boldsymbol{\Phi}_0(\Delta t/2)$. With the resulting basis $\mathcal{V}_m$ and projected matrix $\mathcal{H}_m$, one can approximate the individual matrix exponentials according to

$$e^{(\Delta t - \tau_i)\mathcal{H}} \boldsymbol{\Phi}_0(\tau_i) \approx \mathcal{V}_m\, e^{(\Delta t - \tau_i)\mathcal{H}_m} \mathcal{V}_m^{\mathrm{T}} \boldsymbol{\Phi}_0(\tau_i).$$

Inserting this approximation into Eq. (6.10), one obtains the full time-stepping procedure

$$\mathbf{q}(t_0 + \Delta t) = \mathbf{q}(t_0) + V_m \sum_{i=1}^n w_i\, e^{(\Delta t - \tau_i)\mathcal{H}_m} \mathcal{V}_m^{\mathrm{T}} \boldsymbol{\Phi}_0(\tau_i). \tag{6.11}$$

The computational cost of the inclusion of sources is now dominated by one additional matrix-vector-product in $\boldsymbol{\Phi}_0$ and n corresponding products with $\mathcal{V}_m^{\mathrm{T}}$. If one takes into account that $\mathbf{j}$ is usually an extremely sparse vector, the latter contribution can be reduced to only one full multiplication with $\mathcal{V}_m^{\mathrm{T}}$ and n sparse matrix-vector-products $\mathcal{V}_m^{\mathrm{T}}\mathbf{j}(\tau_i)$.

The stability and accuracy of this procedure strongly depends on the chosen quadrature scheme. A criterion for the choice of a stable quadrature is available [58], but a full analysis of the influence of the quadrature scheme on the accuracy of the result remains an open issue. In practice, we use a 10-point closed Newton-Cotes quadrature. Because the nodes are equally spaced for this scheme, it is sufficient to only evaluate $\exp\left(\tau_1 H_m\right)$ via Padé approximation and all further matrices can be produced by simple matrix-matrix multiplications. As we will see below, this type of quadrature is rather limiting if large time steps are required. In this case it might be advantageous to either use a composite Newton-Cotes formula or switch to a specialized Gaussian quadrature

[60]. The latter is of particular interest because it should enable us to absorb part of the source, e.g., the carrier wave, into the quadrature. Thus, such an approach has the potential to dramatically improve the accuracy and/or the performance of the method.

Sources via Auxiliary Differential Equations

While the direct integration of the convolution integral as discussed in the previous section is capable of dealing with arbitrary source terms, it can become rather time-consuming and potentially reduces the accuracy. As an alternative, we have developed an approach which still allows us to include almost arbitrary sources at significantly reduced computational cost and without any loss of accuracy or time-stepping capabilities. More precisely, we attempt to model the source term $\mathbf{j}(t)$ through a system of linear ADEs (cf. Ch. 3). Including such a system into $\mathbf{q}$ and $\mathcal{H}$ allows us to remove $\mathbf{j}(t)$ from Eq. (6.8) and reduces our problem to the evaluation of a single—albeit slightly more complex—matrix exponential.

The general principle of this scheme may best be demonstrated through the example of a simple time-harmonic electric current source $\vec{j}_E(t) = \vec{j}_0 j(t)$, where $j(t) = \sin(\omega t)$. Obviously, $j(t)$ is a solution of the (auxiliary) harmonic oscillator equation

$$\partial^2 j(t)/\partial t^2 = -\omega^2 j(t)$$

with appropriate initial conditions. Since this differential equation is of second order in time, we first have to reduce it to a system of first order equations. This is facilitated by introducing $k(t) = \partial j(t)/\partial t$. In total, we obtain the complete scheme

$$\frac{\partial}{\partial t} \begin{pmatrix} \vec{E} \\ \vec{H} \\ j \\ k \end{pmatrix} = \begin{pmatrix} 0 & \frac{1}{\epsilon}\nabla\times & \vec{j}_0 & 0 \\ -\frac{1}{\mu}\nabla\times & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -\omega^2 & 0 \end{pmatrix} \begin{pmatrix} \vec{E} \\ \vec{H} \\ j \\ k \end{pmatrix}. \tag{6.12}$$

The extension of this procedure to more complicated sources exploits two basic properties of our solver. First, we carry out a time-stepping procedure which means that we can efficiently approximate our sources by piecewise defined functions, as long as each piece is sufficiently smooth and much longer than our timestep Δt. Second, we can easily add any homogeneous linear system of differential equations to our problem under consideration. In order to simplify the notational burden, we will consider a single linear ADE of order n that should model the source $j(t)$

$$a_n j^{(n)}(t) + \cdots + a_2 j''(t) + a_1 j'(t) + a_0 j(t) = 0. \tag{6.13}$$

We assume real coefficients a_i, $i = 1, \ldots, n$ which makes it straightforward to transform this equation into a system of real first-order ordinary differential equations as required by our solver. The general solution of Eq. (6.13) is

$$j(t) = \sum_{i=1}^{k} \left[\sum_{j=0}^{l_i-1} \left(c_j^{(i)} t^j \right) e^{\lambda_i t} \right], \tag{6.14}$$

where k is the number of linearly independent roots λ_i with multiplicity l_i ($i = 1, \ldots, k$) of the characteristic equation of Eq. (6.13). In addition, the coefficients $c_j^{(i)}$ ($j = 1, \ldots, l_i$) are determined by the initial conditions.

This result shows that we can directly model sources with a polynomial envelope and an underlying harmonic oscillation (for imaginary λ_i). Exactly this feature will be exploited in the following to model general sources by piecewise approximation of the exact envelope through a polynomial of order n. To do so, we first need to find the polynomial $p_n(t) = \sum_{k=0}^{n} c_k t^k$ of order n which approximates the envelope of our source on a given interval $[t_a, t_b]$. The optimal (in a least-square sense) coefficients c_i can be found by an expansion of the target source into Legendre polynomials [43]. This procedure requires us to solve the equations

$$\sum_{k=0}^{n} c_k \int_{t_a}^{t_b} t^{j+k} \, dt = \int_{t_a}^{t_b} j(t) t^j \, dt, \quad j = 0 \ldots n \tag{6.15}$$

for the coefficients c_i.

Once we have obtained the coefficients, we go backwards and construct the correct ADE which has the desired polynomial as a solution. For a carrier frequency ω_0, we require the complex pair $\lambda_{\pm} = \pm i\omega_0$ as roots of the characteristic equation. Further, each root has to be $(n+1)$-fold degenerate in order to allow for a polynomial envelope of order n. Thus we create the characteristic equation as

$$(\lambda + i\omega_0)^{n+1} (\lambda - i\omega_0)^{n+1} = 0.$$

From this, we can directly generate the required system of ADEs and then use the initial conditions in order to determine the coefficients c_i.

We now illustrate this somewhat abstract procedure through the explicit construction of a system of ADEs for the source $j(t)$ with a Gaussian temporal profile which we have used in Sec. 4.4.3, i.e.,

$$j(t) = e^{-t^2/(2\sigma^2)} \cos(\omega_0 t). \tag{6.16}$$

First, we limit our source to the interval $[-5\sigma, 5\sigma]$. For t outside this region, we set $j(t)$ to zero. Next, we split the interval into two parts, the ramp-up part $[-5\sigma, 0]$ and the ramp-down part $[0, 5\sigma]$. We restrict the discussion to the ramp-up part, the ramp-down being perfectly analogous. In Fig. 6.2 we demonstrate that an 11th-order polynomial represents an excellent approximation to the envelope of this source. The corresponding coefficients c_i are obtained by solving Eqs. (6.15) and are listed in Tab. 6.1.

Finally, after having set up the characteristic function for $n = 11$, we obtain the

Coefficient	Value
c_0	0.00002738064
c_1	$-0.00084243573/\sigma$
c_2	$0.00763879040/\sigma^2$
c_3	$-0.02842436916/\sigma^3$
c_4	$0.05676807528/\sigma^4$
c_5	$-0.06672686731/\sigma^5$
c_6	$0.04935061201/\sigma^6$
c_7	$-0.02332427648/\sigma^7$
c_8	$0.00703431090/\sigma^8$
c_9	$-0.00128177207/\sigma^9$
c_{10}	$0.00012671339/\sigma^{10}$
c_{11}	$-5.18632981473 \times 10^{-6}\sigma^{11}$

Table 6.1: *Coefficients for an 11th-order polynomial $p_{11} = \sum_{k=0}^{11} c_k t^k$ that approximates the ramp-up part of a pulse with Gaussian temporal profile.*

following system of 24 first-order ADEs

$$
\begin{aligned}
\frac{\partial}{\partial t}j(t) &= j_1(t), \\
\frac{\partial}{\partial t}j_i(t) &= j_{i+1}(t), \quad i = 1, \ldots, 22, \\
\frac{\partial}{\partial t}j_{23}(t) &= -12\omega_0^2 j_{22} - 66\omega_0^4 j_{20} - 220\omega_0^6 j_{18} - 495\omega_0^8 j_{16} - 792\omega_0^{10} j_{14} - 924\omega_0^{12} j_{12} \\
&\quad -792\omega_0^{14} j_{10} - 495\omega_0^{16} j_8 - 220\omega_0^{18} j_6 - 66\omega_0^{20} j_4 - 12\omega_0^{22} j_2 - 12\omega_0^{24} j.
\end{aligned}
$$

What remains is to determine the initial conditions, i.e., $j(t = 0)$ and $j_i(t = 0)$. This is facilitated by successively differentiating our solution

$$
j(t) = \left(\sum_{i=0}^{11} c_i t^i \right) \cos(\omega_0 t)
$$

and inserting the coefficients from Tab. 6.1. Therefore, in order to fully describe the ramping-up and -down of a Gaussian pulse, we need to add $2 \times 24 = 48$ ADEs to our basic solver. In more extreme cases, the number of resulting ADEs may go into the hundreds. In view of the fact that modeling a typical nanophotonic system requires at least some $10^5 - 10^6$ unknowns, the additional computational overhead incurred can then still be neglected.
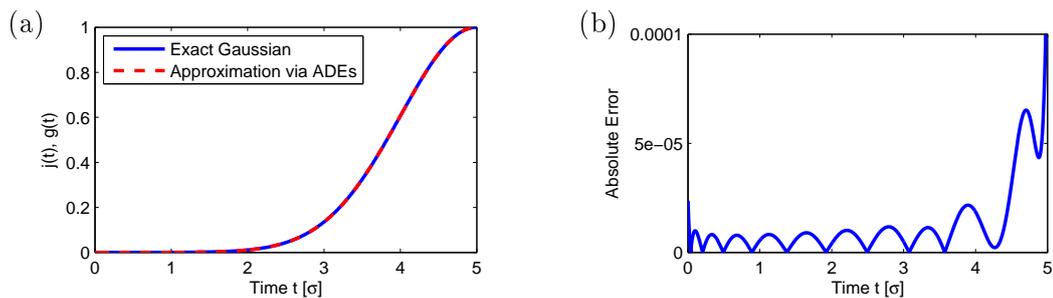
Figure 6.2: *Quality of the approximation of a source with Gaussian temporal profile via an 11th-order polynomial. (a) Shape of the ramp-up part of the envelope, (b) Error of the polynomial approximation to the Gaussian profiles, Eq. (6.16).*

Thus, we have a general scheme for the inclusion of almost arbitrary source terms into the Maxwell equations without ever evaluating the convolution integral in Eq. (6.9). Further, since the source is fully described by a system of differential equations, we never have to sample the corresponding source function explicitly. In turn, this implies that our time steps are not limited by the sampling theorem's apparent requirement of sufficiently over-sampling the source terms.

6.2 Numerical Experiments

Finally, we provide details about the solver's performance and compare it with the standard FDTD approach. As discussed in the introduction to this chapter, the most significant gains can be expected for higher-order discretizations. Therefore, for all further computations in this chapter, we employ a Yee-grid with a 4th-order accurate discretization as discussed in Sec. 4.2.3.

6.2.1 One-Dimensional Systems

We start with a simple 1D test setup consisting of a system of length $L = 20$ (dimensionless units), the right half of which is filled with a material of refractive index $n = 1.5$. The left side is vacuum (refractive index $n = 1$). A Gaussian pulse is prepared on the left side and is allowed to propagate towards the interface through the initial conditions $E(x, t) = \exp\left(-(x - x_0 - t)^2/\sigma_x^2\right)$ and $H(x, t) = -E(x, t)$ ($x_0 = 5$ and $\sigma_x = 1$). For FDTD, which works in a time-staggered fashion, we initialize the system with $E(x, 0)$ and $H(x, \Delta t/2)$, whereas the Krylov subspace method assumes all initial conditions at the same time. When the pulse hits the interface it splits into reflected and transmitted parts. After the computational time $T = 12$ has elapsed, we compare the electric field of the transmitted pulse with the values of the corresponding analytical solution. In Fig. 6.3 we display the relative error of the computations as a function of the Courant number s=$\Delta t/\Delta x$. As expected, once the CFL stability criterion is
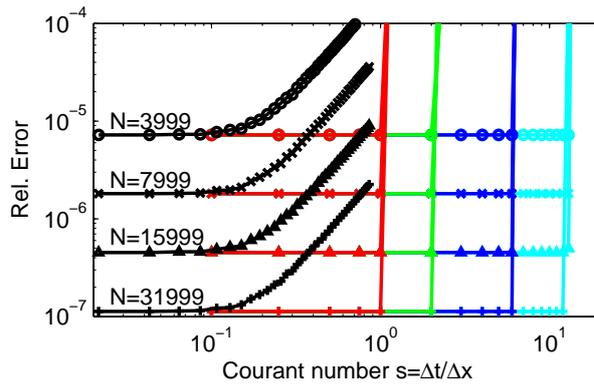
Figure 6.3: *Comparison of the standard FDTD method with the Krylov-subspace approach in a one-dimensional system. The symbols indicate the number of total unknowns employed in the computation. We have used N_{tot}=3999 (circles), N_{tot}=7999 (crosses), N_{tot}=16999 (triangles), and N_{tot}=31999 (pluses) grid points. The line colors correspond to the methods: FDTD (black) and the Krylov-subspace approach with subspace dimensions $m = 4$ (red), $m = 8$ (green), $m = 16$ (blue), and $m = 32$ (cyan), respectively.*

satisfied ($s \leq \frac{6}{7}$), the error for the FDTD results decrease quadratically with decreasing time steps, until the curve flattens out and approaches the error bound set by the accuracy of the spatial discretization. In contrast, for the Krylov-subspace method the relative error is of the order one unless the time step is sufficiently small for the method to converge. When this happens, the error drops extremely fast to the error bound set by the spatial discretization. The value of Δt at which this drop happens depends on the subspace dimension, but the decay itself is always almost "instantaneous". These results demonstrate that within the Krylov-subspace approach we can use more than two orders of magnitude larger time steps as compared to FDTD. We would like to emphasize that this behavior is a generic feature of the Krylov-subspace approach and does not depend on the choice of spatial discretization.

6.2.2 Two-Dimensional Systems

As a first comparison in 2D, we choose the test setup depicted in Fig. 6.4 with a total system size of 8×4. As initial conditions, we prepare a localized pulse in the (x,y)-plane that propagates into the positive x-direction by

$$H_x(x, y, t = 0) = 0,$$
$$H_y(x, y, t = 0) = -E_z(x, y, 0),$$
$$E_z(x, y, t = 0) = \cos(k_0(x - x_0))e^{-(x-x_0)^{10}/\sigma_x^{10}}e^{-y^2/\sigma_y^2}.$$

The values of the parameters are $x_0 = -1.25$, $k_0 = 2\pi$, $\sigma_x = 1$, and $\sigma_y = 0.5$. In the right part of the system, we have positioned two cylinders with radius $r = 0.5$

(a)

(b)
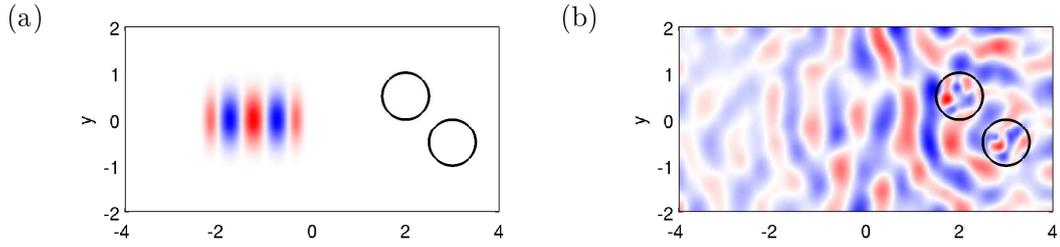
Figure 6.4: *Snapshots of the electric field E_z for the 2D test setup at (a) $T = 0$ and (b) $T = 10$. The circles indicate the positions and sizes of cylinders filled with refractive index $n = 2$.*

(a)

(b)
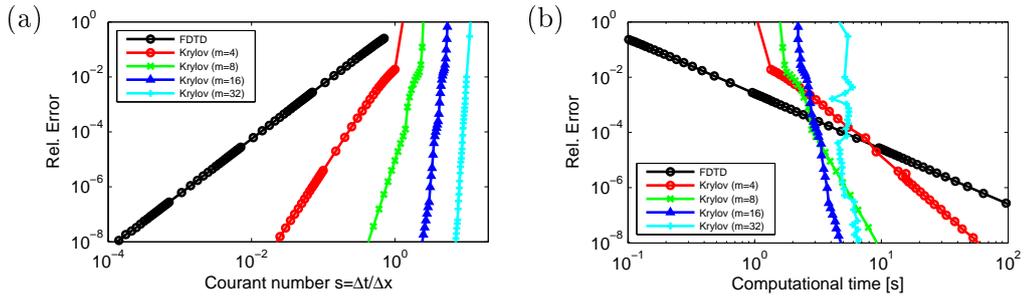
Figure 6.5: *A comparison between standard FDTD and the Krylov-subspace method for the 2D test system depicted in Fig. 6.4. (a) displays the error as a function of the Courant number, while (b) shows the error as a function of the actual computational time on a computer with an AMD Athlon64 3500+ processor.*

and refractive index $n = 2$. Their centers are, respectively, located at $(2, 0.5)^{\mathrm{T}}$ and $(3, -0.5)^{\mathrm{T}}$. In order to overcome the problem that FDTD requires time-staggered initial conditions, we first use the Krylov subspace method with $m = 32$ and at least 10 time steps, to obtain the initial values for $H_x(x, y, t = \Delta t/2)$ and $H_y(x, y, t = \Delta t/2)$ from the initial value of the electric field.

With these initial conditions, we let the system evolve until $T = 10$. In Fig. 6.4 we display the electric field E_z before and at the end of the computation. The Yee-grid consists of 160×80 cells, where the electric field E_z is defined in the cell center while the magnetic fields, H_x and H_y, reside on the cell vertices. We have implemented perfectly reflecting boundary conditions, by requiring the magnetic field at the system boundaries to be zero (perfect magnetic conductor). We double-checked this procedure by using a complementary FDTD calculation with very small time steps and both results agree up to machine precision. Since this problem is not amenable to an analytical solution, we use as a reference-solution the Krylov-subspace method with $m = 32$ and $\Delta t = 0.01\Delta x$. This reference solution is (up to machine precision) identical to a solution obtained from FDTD with $\Delta t = 10^{-6}\Delta x$. In Fig. 6.5, we compare the accuracy of the two methods as a function of both the Courant number $s = \Delta t/\Delta x$ and the actual

(a)

(b)
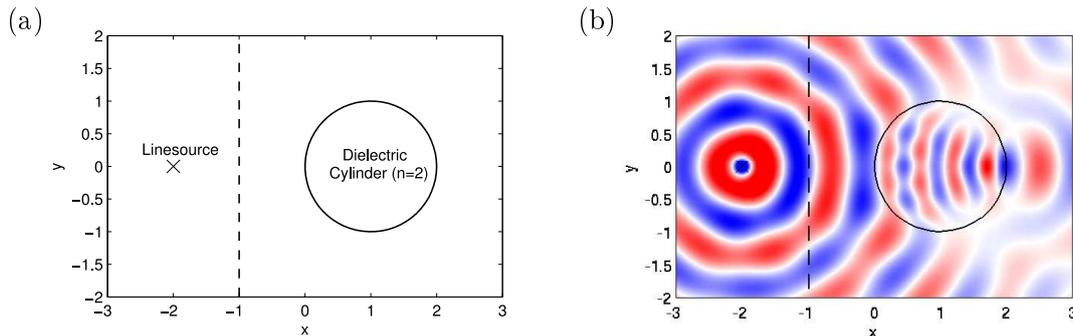
Figure 6.6: *Schematic setup (a) and reference solution (b) of the second (open) 2D test system. A dielectric cylinder (radius $r = 1$, index of refraction $n = 2$) is centered at $(1, 0)^{\mathrm{T}}$ and is illuminated by an electric line source (frequency $\omega = 2\pi$) located at $(-2, 0)^{\mathrm{T}}$. In both panels, we delineate a dashed line. The entire computational domain to the right of this line is used for comparing the results of the Krylov-subspace approach.*

computational time. As expected, for FDTD the error drops quadratically, while for the Krylov-subspace method the error decays as Δt^m. Looking at the computational times, obtained on a desktop computer with a AMD Athlon64 3500+ processor (see Fig. 6.5(b)), the Krylov-subspace method becomes advantageous for a relative error below 10^{-4}. For very high accuracy calculations, the Krylov-subspace approach can be two to three orders of magnitude more efficient than standard FDTD. From Fig. 6.5(b) we also see that increasing m from 16 to 32 results in a slight performance loss, which largely stems from the overhead of the exponential of H_m and the much larger V_m. Thus, for optimal performance one must not choose m too large. Depending on the required accuracy and the available memory, we propose to work with Krylov-subspace dimensions ranging from m from 4 to 12. Concerning the memory usage, the Krylov-subspace approach generally requires more memory than standard FDTD. For typical applications where the entire basis is kept in the memory, the Krylov-subspace approach requires storage for a total of $(m + 1)N_{\mathrm{tot}}$ unknowns. Therefore, accuracy and long time-stepping capability have to be balanced against available memory.

6.2.3 Testing of the Sources

To demonstrate the capability of the Krylov-subspace approach to model open systems and sources, we choose a slightly different system whose setup is sketched in Fig. 6.6. The total system size is 6×4 (dimensionless units) and it is discretized by 301×201 grid cells. A dielectric cylinder (radius $r = 1$, refractive index $n = 2$) is centered at $(1, 0)^{\mathrm{T}}$ and an electric line source with frequency $\omega = 2\pi$ is positioned at $(-2, 0)^{\mathrm{T}}$. Since we are simulating an open system, we have added CFS-PMLs with a thickness of ten grid cells to the entire computational domain. To obtain a reference solution for the electric field, we have employed the (for such simple systems numerically exact)
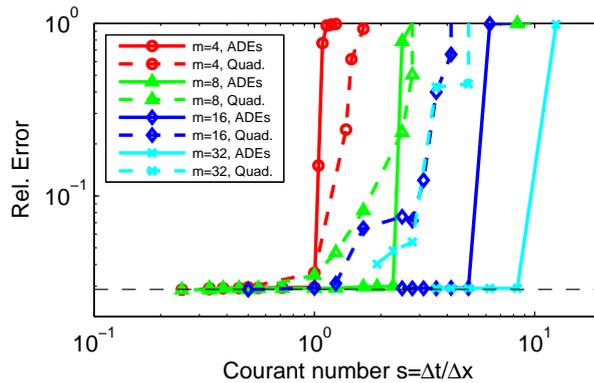
Figure 6.7: *Convergence study of the Krylov-subspace method for an open system that includes a source. The plot displays the relative deviations from the reference solution as a function of the Courant number. Two different implementations of the sources have been utilized: A harmonic source via ADEs (solid) and an approximation via a quadrature formula for the convolution integral (dashed). The dotted line represents the error bound set by the spatial discretization. This error bound has been obtained by a fourth-order FDTD-calculation with $s = 0.1$.*

multiple multipole expansion technique [61]. Since this reference solution represents the steady-state of the system, we have to run our time-domain simulations sufficiently long so that transients have died out. In practice, we run the simulations for $T = 250$ and compare the computed field in the region of the computational domain described by $x \geq -1$ with that of the reference solution.

We model this system with the source realized by ADEs (see Eq. (6.12)) and, alternatively, by a 10-point closed Newton-Cotes quadrature of the convolution integral formulation (see Sec. 6.1.2). The results in Fig. 6.7 indicate that sources via ADEs work extremely well. In addition, these results highlight the capability of the Krylov-subspace approach of doing very large time steps even in the presence of sources. Note that in the case of $m = 32$ the method converges with less than 3 time steps per oscillation period of the source. These results also suggest that the inclusion of PMLs does not have any influence on the stability or accuracy of the Krylov-subspace approach.

6.3 Conclusions on the Choice of Time-Integration

In total, we have discussed three different time-integration schemes in this thesis: The finite-difference scheme (FDTD), a low-storage Runge-Kutta method (DG) and now the Krylov-subspace approach. All three methods have their own advantages and disadvantages, which we tried to summarize in Tab. 6.2. When looking for a suitable method, it almost always comes down to a trade-off between available memory, performance and accuracy. Generally speaking, the simpler methods require less memory but they also perform worse in terms of computational time or accuracy. Hence, one

should pick an appropriate scheme on the basis of the problem at hand. For simpler problems with low requirements on accuracy, the integration via finite differences might be sufficient. In contrast, for challenging but not memory limited problems, the Krylov subspace approach is a viable alternative. Between these two extremes, the low-storage Runge-Kutta scheme presents a good compromise for many cases.

As a final note, it should be stated that there exists a wealth of other integration methods and new schemes are still developed regularly. Our selection of methods is by no means comprehensive and certainly there are other methods which are more suitable for one problem or another. While it is doubtful that there exists an ultimate method, we still expect that further research in the field of time integration will lead to significant performance gains.

	Finite Differences (FDTD)	Low-Storage Runge-Kutta	Krylov-subspace
Accuracy	Second-order accurate	Can be constructed for different orders, coefficients known up to fourth order	Order can be chosen arbitrarily
Stability	Conditionally stable (CFS limit known analytically)	Conditionally stable (CFS limit only known empirically)	Unconditionally stable, practically limited by accuracy
Performance	Only one matrix-vector product per step, no overhead	Number of matrix-vector products varies, usually 5-6 for a fourth-order scheme, little overhead	$m+1$ matrix-vector products per timestep, overhead increases strongly with m
Memory Requirements	Storage for the N_{tot} unknowns, due to leap-frogging	Storage of $2N_{\text{tot}}$, if the matrix-vector product can by directly added to a given vector	Storage of $(m+1)N_{\text{tot}}$ plus additional storage for a $m \times m$-matrix
Flexibility	General, but massive problems with stiff and/or highly oscillatory systems.	Very general, might have problems with very stiff systems	Limited to linear problems (non-linear extensions possible [55]), sources require additional effort, advantages for stiff and highly oscillatory problems

Table 6.2: *Brief summary of the advantages and disadvantages of the three different time-integration schemes discussed in this thesis.*

7

Chapter 7

Optical Micro-Resonators

Optical resonators have come a long way since the pioneering work of Fabry and Perot in 1899 [62]. Today, different kinds of electromagnetic resonators have found applications as filters and antennas from radio frequencies to the visible spectrum. Not to forget, resonators are an essential part of practically every laser setup. A common feature of all resonator structures is their ability to selectively modify certain optical properties within narrow spectral ranges. Recently, progress in nano-fabrication technology has allowed to manufacture optical resonators with dimensions of only a few optical wavelengths. The interest in such microresonators ranges from applications as sensors [8] to fundamental questions of light-matter interaction [63]. Typically, the resonators must be tailored specifically for a given task. For these design purposes, accurate and fast numerical calculations are indispensable.

This chapter starts with a very brief introduction to the theory of optical resonators. Next, we will discuss a special type of optical resonators, namely dielectric cylinders which can be treated analytically. These exact results will then be used to evaluate and compare numerical results obtained with FDTD and the DG method. Finally, we apply both methods to realistic cases of integrated disk and ring resonators.

7.1 Fundamentals of Optical Micro-Resonators

Optical resonators or microcavities are essentially defined by their capability to trap light for a certain time in a restricted volume. The underlying mechanism for the confinement is typically either total internal reflection or distributed Bragg reflection. Starting from the linear Maxwell equations in the conservation form (2.11), we can apply a Fourier transform in time to obtain the eigenproblem

$$\mathcal{Q}^{-1} \left[\mathcal{A}_x \partial_x + \mathcal{A}_y \partial_y + \mathcal{A}_z \partial_z \right] \mathbf{q} = i\omega \mathbf{q}. \tag{7.1}$$

Here, the matrix $\mathcal{Q}$ contains the material distribution (see Sec. 2.2) and therefore defines the resonator setup. As solutions to Eq. (7.1), one finds a set of eigenfrequencies ω_j and the corresponding eigenvectors $\mathbf{q}_j$. To qualify as a proper resonator mode, we

expect the fields to be strongly localized in the vicinity of the resonator. Further, the eigenfrequencies should be discrete and well separated. In general, the frequencies ω_j will not be purely real, but instead they contain a negative imaginary part. This imaginary part describes the decay of a mode in time.

To distinguish between the real and the imaginary part of ω_j, we define the real resonance frequency ν_j as

$$\nu_j = \frac{\text{Re}\left[\omega_j\right]}{2\pi},$$

and the real quality factor or Q factor as

$$Q_j = -\frac{\text{Re}\left[\omega_j\right]}{2\text{Im}\left[\omega_j\right]}.$$

Q_j is proportional to the ratio between the stored energy and the energy dissipated in one oscillation cycle [20]. Therefore, it is a measure for the lifetime of the corresponding mode. The Q factor is determined by a number of different contributions

$$\frac{1}{Q} = \frac{1}{Q_{\text{rad}}} + \frac{1}{Q_{\text{surf}}} + \frac{1}{Q_{\text{mat}}}.$$

Here, Q_{rad} denotes the radiative losses due to curvature, Q_{surf} describes scattering losses caused by imperfections of the surface and Q_{mat} contains the intrinsic material losses. In the following, we will neglect Q_{mat} and concentrate on Q_{rad} and particularly on the effect of an artificial Q_{surf}, which is introduced by the numerical discretization of the surface.

7.1.1 Numerical Extraction of the Resonance Frequencies

As discussed above, one needs to solve the eigenproblem (7.1) in order to obtain the resonance frequencies and quality factors of a given resonator. While it is possible to solve (7.1) directly, here, we will exploit the time-domain solvers discussed in the previous chapters.

As a first step, we have to excite the modes of interest. In some cases, this can be achieved via initial conditions, but it usually is more convenient to employ one or multiple point sources. Tuning the spectrum of these sources allows for selective excitation of a certain set of modes. Once the modes are excited, we turn off the sources and start to record the fields at a point $\vec{r}_0$ in the resonator. In principle, the recording point can be chosen arbitrarily, as long as the mode does not exhibit a node at this position. Since the eigenmodes form a complete basis of our system, we expect to observe a superposition of the different resonator modes

$$\mathbf{q}(\vec{r}_0, t) = \sum_j \mathbf{q}_j(\vec{r}_0)e^{-i\omega_j t} = \sum_j \mathbf{q}_j(\vec{r}_0)e^{-2\pi\nu_j\left(i+\frac{1}{2Q_j}\right)t}.$$

To extract the frequencies, we apply a Fourier transform to the recorded data and obtain

$$\breve{\mathbf{q}}(\vec{r}_0, \nu) = \int\limits_0^\infty dt\, \mathbf{q}(\vec{r}_0, t) e^{-i2\pi\nu t} = \sum_j \mathbf{q}_j(\vec{r}_0) \frac{iQ_j}{2\pi Q_j\left(\nu - \nu_j\right) + i\pi\nu_j}. \qquad (7.2)$$

It should be noted that the Fourier integral's lower limit is $t = 0$, because we assume the recording to start at this time. Considering only a single resonance and taking the modulus squared of Eq. 7.2 then yields

$$\left|\breve{\mathbf{q}}_j(\vec{r}_0, \nu)\right|^2 = \frac{1}{\pi^2} \frac{1}{4\left(\nu - \nu_j\right)^2 + \left(\frac{\nu_j}{Q_j}\right)^2},$$

which we recognize as a Lorentzian curve. If we assume all resonances in our system to be well separated, then the modulus squared of Eq. (7.2) will be a sum of Lorentzian line shapes. By plotting this data, we can identify the resonance frequencies and Q factors via the peak positions and peak widths, respectively.

Unfortunately, above procedure has a number of serious limitations. First, we need to record the data until all fields have decayed sufficiently. Otherwise, we will introduce spurious oscillations due to the Fourier transform of the cut-off. Second, it is far from trivial to distinguish the different modes in a plot, if the Q factors are not sufficiently large or if the frequencies are not well separated. To overcome these problems, a number of different analysis tools were developed. Most of them do not execute a Fourier transform but rather try to directly fit a sum of exponentially decaying functions to the recorded time data. One of the most effective methods in this field is known as the filter diagonalization method (FDM) [64]. It allows to detect a large number of modes from a relatively small amount of data. In the following, we will employ the library `harminv` [65], which is a very convenient and freely available implementation of the FDM.

7.2 Analytic Test Cases

For certain resonator geometries, it is possible to find the resonance frequencies analytically. A particularly elegant method was introduced by Gustave Mie in 1908 [66], which allows the exact treatment of cylindrical and spherical geometries. Here, we will use this technique to obtain reference solutions for comparison with our numerical results. A complete discussion of Mie's theory is far beyond the scope of this thesis. The interested reader is referred to a number of books [67–69] for the details of the analytic calculations.

7.2.1 Infinite Cylinders under Normal Incidence

An infinite cylinder with light propagation restricted to the plane perpendicular to the cylinder axis can be described as a two-dimensional system. As we have seen in

m	l = 1		l = 2	
	ν	**Q**	ν	**Q**
17	1.955	560.1	2.364	59.62
18	2.065	829.5	2.479	73.33
19	2.175	1236	2.594	90.99

Table 7.1: *An exemplary set of resonance frequencies and their respective Q-factors for a cylinder in TM polarization with radius $r = 1$ and refractive index $n = 1.59$ (cf. Fig. 7.1).*
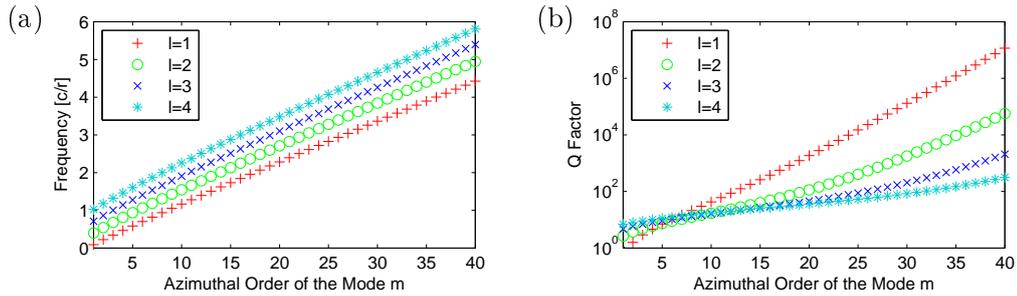
Figure 7.1: *Analytic results for the modes of a cylinder with $n = 1.59$ and radius $r = 1$ in TM polarization. Different symbols/colors denote different radial order l. (a) shows the resonance frequencies, while (b) contains the corresponding Q factors.*

Sec. 2.3, TM and TE polarizations decouple for this case and we expect two sets of independent solutions. Assuming a cylinder with refractive index $n = \sqrt{\epsilon}$ and radius r, the complex resonance frequencies ω_j can be found as the zeros of the following equations [69]:

$$\text{TM-Polarization:} \quad J_m(n\omega)H_m^{(1)'}(\omega r) - nJ_m'(n\omega)H_m^{(1)}(\omega r) = 0, \tag{7.3a}$$

$$\text{TE-Polarization:} \quad nJ_m(n\omega)H_m^{(1)'}(\omega r) - J_m'(n\omega)H_m^{(1)}(\omega r) = 0. \tag{7.3b}$$

Here, J_m are the Bessel functions, while $H_m^{(1)}$ denote Hankel functions of the first kind. Further, $'$ denotes the derivative with respect to the argument and the index m signifies the azimuthal order of the resonance. For each index m, multiple roots can be found, which are then labeled by the index l. Therefore, we can uniquely identify a cylinder mode by the tuple (m,l).

As an example, we pick a cylinder in TM polarization with radius $r = 1$ and refractive index $n = 1.59$. A few selected resonances and their corresponding Q factors are listed in Tab. 7.1 .

From the results depicted in Fig. 7.1, we observe that the resonance frequencies grow linearly with the azimuthal order m. Modes of higher radial order l also have higher
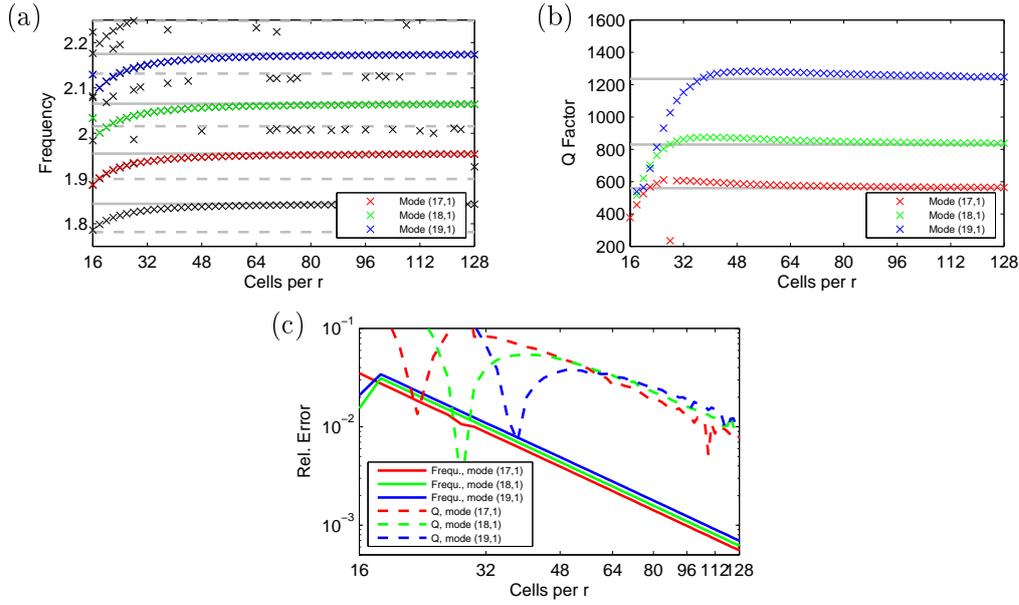
Figure 7.2: *Resonator modes found from FDTD calculations for different spatial discretizations. Panels (a) and (b) contain the frequencies and Q factors, respectively. Analytic solutions are indicated by horizontal lines for l=1 (solid) and l=2 (dashed). The modes (17,1), (18,1) and (19,1) are highlighted and their corresponding relative error is plotted in panel (c).*

frequencies. In contrast, the Q-factors grow exponentially with m but typically, they are maximal for $l = 1$. Therefore, if a large Q is desired, one should pick a mode of high azimuthal order m but with minimal radial order $l = 1$.

Calculations with FDTD

For the calculations with FDTD, we model a system of extent 3×3 (in dimensionless units) with the cylinder of radius $r = 1$ centered in it. Around the system, we add 10 cells of PMLs in each direction. For excitation, we employ an electric dipole source, oriented in z-direction at position $\vec{r}_{\text{source}} = (0.8, 0)^{\text{T}}$. The time dependence of our source is given by

$$j(t) = \sin(2\pi\nu_0(t - t_0))e^{-\frac{(t-t_0)^2}{2\sigma^2}},$$

where $\nu_0 = 2$, $\sigma = 2$ and $t_0 = 5\sigma$. These parameters ensure a sufficiently broad spectrum to excite modes with frequencies between $\nu = 1.5$ and $\nu = 2.5$.

After the source has decayed sufficiently ($T = 2t_0$), we start recording the E_z field component at the position of the source for 1000 time units. The resonances are then extracted via `harminv`. The results for different spatial discretizations are depicted in Fig. 7.2. From Fig. 7.2(a), we observe that `harminv` sometimes reports modes, which can not be identified as modes with $l = 1$. Whether these modes are of higher
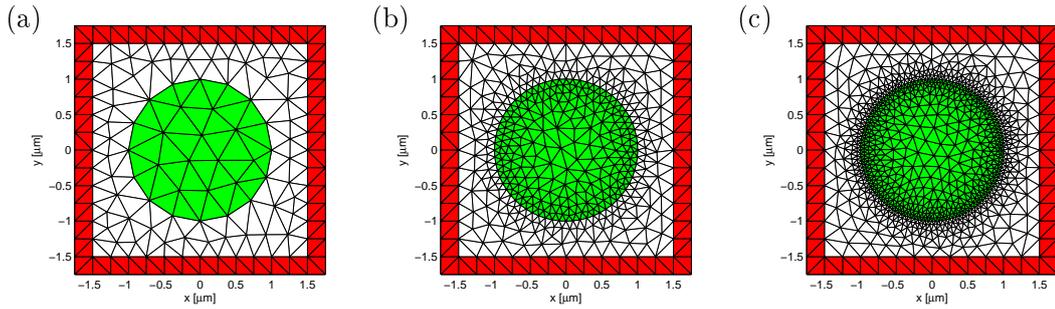
Figure 7.3: *Cylinder meshes with different refinement: (a) $\alpha = 1$, (b) $\alpha = 5$, (c) $\alpha = 10$. The green areas denote regions with refractive index $n = 1.59$, while the red shaded triangles belong to the PMLs.*

radial order or spurious is difficult to decide. Since they lie reasonably close to analytic resonances with $l = 2$, we attribute them to weak contributions from these modes.

Concerning the first-order modes, from the relative error plotted in Fig. 7.2(c), we find the expected quadratic convergence for the resonance frequencies. We also observe that the error drops below 1% at approximately 32 cells per radius. If we relate this to the wavelength in the material, given by $\lambda_{\mathrm{mat}} = 1/(n\nu_0) = 0.314$, we recover the rule of thumb, that 10 cells per wavelength are sufficient for an error of approximately 1%. Considering the Q factor however, significantly more cells are required to reach the same level of accuracy. As can be seen from Fig. 7.2(c), even 128 cells per radius are barely sufficient. For realistic calculations, this behavior poses a significant problem, since the optimization of the Q factor often is the main objective.

Calculations with the DG Method

To compare the FDTD results with DG calculations, we first generate a mesh with an edge length of $h_0 = 0.5$. Instead of a uniform refinement, we only improve the discretization of the cylinder, by adding more triangles at the interface. In particular, we introduce a refinement factor α which denotes the ratio between h_0 and the smallest edge of a triangle at the interface. A number of meshes for $\alpha = 1 \ldots 10$ is generated, some examples are shown in Fig. 7.3. To facilitate a fair comparison with the FDTD results, we introduce the average number of gridpoints per radius $n_{\mathrm{avrg}} = \sqrt{N_{\mathrm{Tri}} N_p}/3$. Here, N_{Tri} is the number of triangles in the inner computational domain (excluding the PMLs). N_p is the number of gridpoints per triangle for a given order N. This number can be seen as a rough equivalent to the cells per radius used in the previous section.

In Fig. 7.4 we plot the relative error of mode 18 (cf. Tab. 7.1) as a function of n_{avrg} for different orders N. Studying the convergence in Fig. 7.4(a), we observe that calculations with various orders all seem to exhibit the same slope of approximately 4. This indicates, that our calculations are still limited by the shape of the cylinder surface. In this case, it obviously is futile to increase the polynomial order. Instead
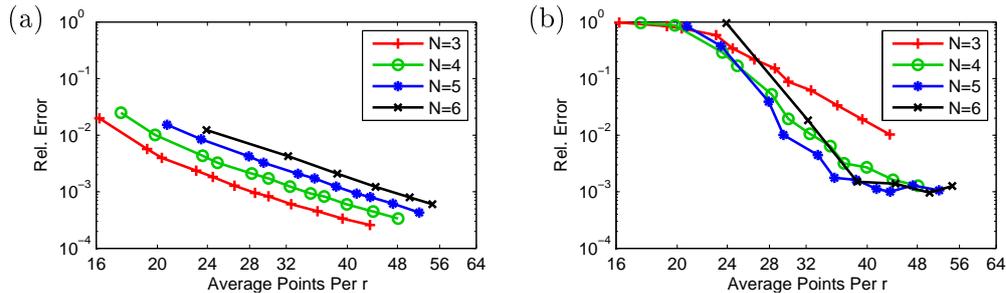
Figure 7.4: *Convergence of the cylinder mode (18,1) for DG calculations. Shown is the relative error of (a) the frequencies and (b) the Q factors as a function of the number of average gridpoints per radius n_{avrg}.*

one should refine the mesh even further, if a higher accuracy is desired. Remarkably, for the Q factors the situation is slightly different. At lower resolutions, the error is reduced with the expected order. Only after it drops below 0.1%, the convergence order reduces to the same as observed for the frequencies. This behavior can be attributed to the dissipative nature of the DG method, which will dominate the error for insufficient resolutions.

However, both the frequencies and the Q-factors can be obtained by the DG method with significantly less grid points as compared to FDTD. For an accurate estimate of the Q factor (relative error below 0.1%), the DG calculations can be up to a factor 100 faster.

7.3 Microcavity Disk and Ring Resonators

Of particular interest for applications are integrated disk- and ring resonators, produced from semiconductor materials like Si or GaAs. Since these structures are usually fabricated by means of electron-beam lithography, the quality of the resonators has profited from the enormous progress of this fabrication technology over the past years. Today, it is possible to produce structures with sizes of tens of microns, where imperfections and surface roughness stay below 5 nm. Here, we picked two typical resonator devices as sketched in Fig. 7.5. The disk resonator in Fig. 7.5(a) is coupled to two waveguides and therefore qualifies as a four-port device. For reasons that will become clear later, this setup is also called an add/drop filter. The second device, pictured in Fig. 7.5(b), only connects to a single waveguide and is therefore called a two-port or all-pass configuration.

7.3.1 Disk Resonators

To allow for direct comparison of our results for the disk resonator with other calculations and experimental measurements, all data in this section will be presented in physical units. For our setup, we choose a disk with radius $r = 2.5\,\mu$m, while the
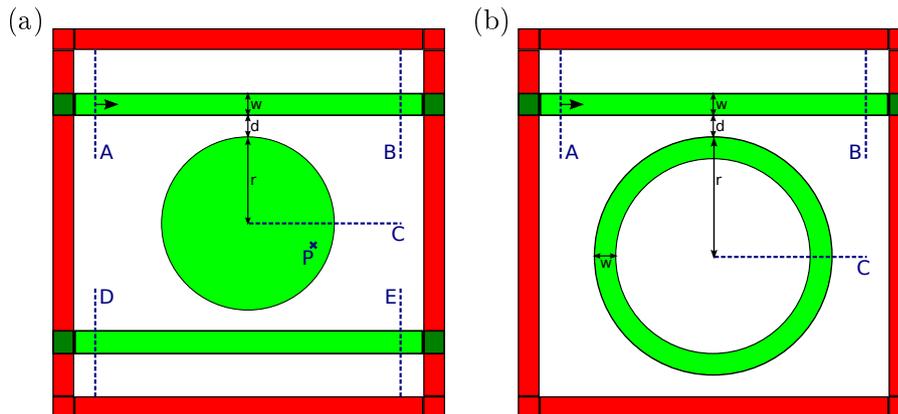
Figure 7.5: *Schematic setup of (a) a disk resonator system and (b) a ring resonator setup. In both cases, the computational domain is surrounded by PMLs. The dashed lines indicate contours through which we measure the flux.*

waveguides have a width of $w = 300$ nm. The distance between disk and waveguides is $d = 232$ nm. The disk and the waveguide are assumed to consist of a material with $\epsilon = 10.24$ ($n = 3.2$). This setup exactly corresponds to the system discussed theoretically and experimentally in Ref. [70].

The aim is to identify the available resonances of the disk resonator at wavelengths around $\lambda_0 = 1.55\,\mu$m. As we have seen from the studies on cylinders in Sec. 7.2, a FDTD calculation will require at least 10 cells per shortest wavelength. Here, the wavelength in the material is given be $\lambda_{\mathrm{mat}} = \lambda_0/n \approx 500$nm. Thus, we would expect to reasonably converged results for a spatial discretization of $\Delta \approx 50$nm. In Fig. 7.6 we show the numerical convergence behavior for a FDTD simulation of our system. Obviously, for this system, the common rule of thumb fails dramatically. A discretization with $\Delta = 50$ nm is far from sufficient and by extrapolation we can estimate a required grid spacing of $\Delta x \approx 10$ nm for the frequency to be converged below 1%. In Ref. [70], the authors employ FDTD calculations with grid spacings of $\Delta = 13.6$ nm, which is roughly consistent with our estimates. Further, they argue that the staircasing serves as a realistic model for the experimental surface roughness. However, by looking closely at the surfaces of structures fabricated with modern microfabrication methods (see e.g. Ref. [71]), this claim seems rather doubtful.

In order to investigate the influence of the surface on the performance of a disk resonator, we employ the DG method. For our studies, we generate four different meshes, "M1" to "M4", as depicted in Fig. 7.7. For each of these meshes, we execute calculations to extract the resonance frequencies of the resonator. In particular, we inject a pulse with a well-defined spectrum in the upper waveguide and let it propagate from left to right. During the simulations, we record the electric field at a point P inside the disk (marked by a cross in Fig. 7.5). Additionally we record the spectral flux through a plane "B" at the right side of the upper waveguide.
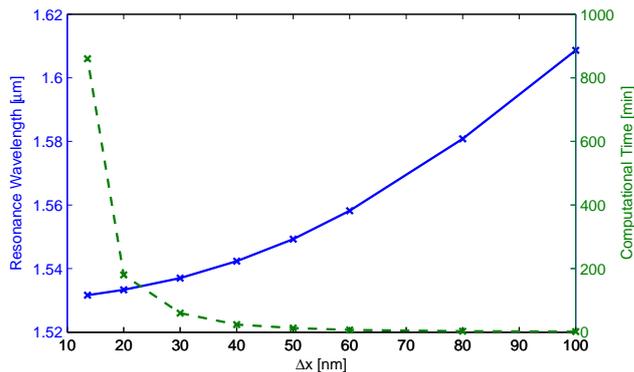
Figure 7.6: *Convergence for a particular mode (m = 28, l = 1) of the disk resonator (solid line). Also plotted is the typical computational time required to obtain the resonance wavelength with a commercial FDTD package on a desktop computer (dashed line).*

To get an overview of the modes around $\lambda_0 \approx 1.55\,\mu$m, we initially excite the system with a broad-band Gaussian pulse. Since most of the resonances are very narrow, a good frequency resolution of 1 nm or better is required. Due to the sampling theorem, this forces us to simulate for rather long times of approximately 334 ps. The normalized spectrum over a broad wavelength region can be found in Fig. 7.8(a). By studying at individual peaks (cf. Fig. 7.8(b)), we observe how the frequencies shift very slightly with the mesh size. Still, from comparison between the data obtained for M3 and M4, we conclude that our results are converged to at least four decimal places. It should be noted that all calculations where conducted with 4th order discretizations.

To obtain more detailed information on the resonances, we also employ a FDM on the data collected at point P inside the disk. From the time-data, plotted in Fig. 7.9(a), we find that even after almost 350 ps (more than 5×10^5 optical cycles), there still is some electric field localized inside the disk. This indicates modes with extremely high Q factors. The recorded timedata is then analyzed with `harminv` as a function of the simulated time. To simplify matters, we concentrate on modes with resonances between $\lambda = 1.5\mu$m and $\lambda = 1.55\mu$m. The resulting wavelengths are plotted in Fig. 7.9(b). The first thing to note is that—except for a few spurious modes for short simulation times—`harminv` consistently finds three modes. By comparison with the data in Fig. 7.8(b), we confirm that the resonance wavelengths agree very well with the results from our Fourier analysis. To complement the data in Fig. 7.9(b), we list a larger number of resonance wavelengths and Q factors in Tab. 7.2. The individual modes are sorted by their azimuthal order m and their radial order l (cf. Sec. 7.2 and Fig. 7.10). As a first observation, we find that the Q factors increase dramatically with increasing radial order. This is in contrast to the analytical predictions made for the modes of a dielectric cylinder (see Sec. 7.2). For comparison, we have calculated the respective resonance frequencies and Q factors for a dielectric cylinder with radius $r = 2.5\mu$m, and
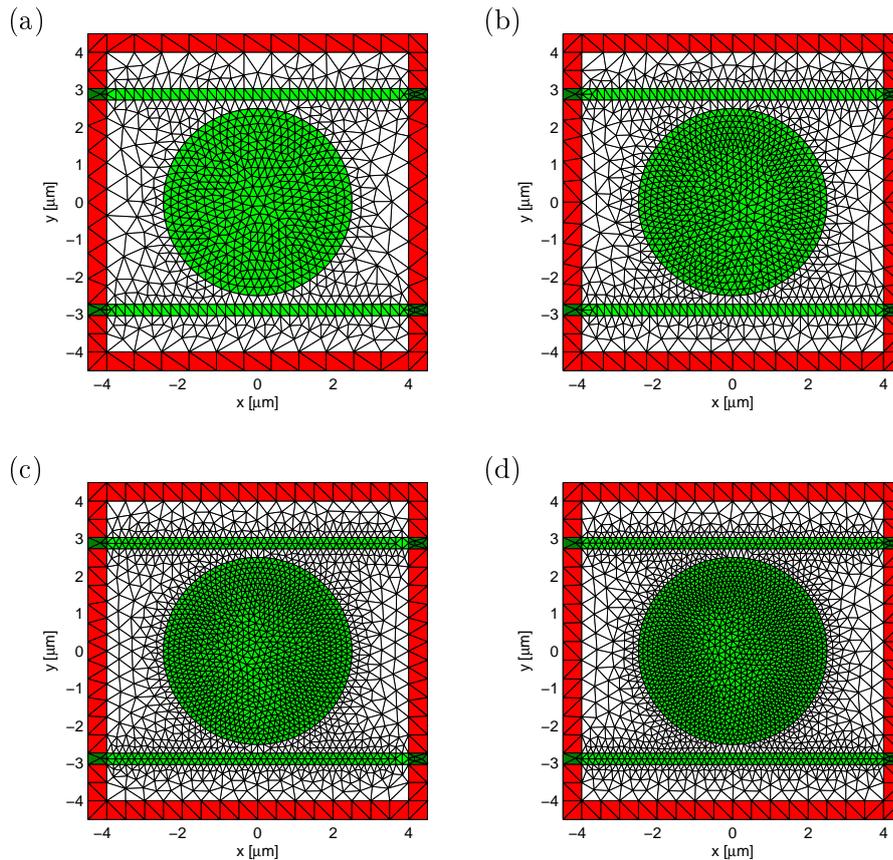
Figure 7.7: *The meshes used for the calculation of the resonance frequencies and spectra. Different colors denote different material properties: air (white), silicon (green) and PMLs (red). In the text, we refer to the meshes as (a) M1, (b) M2, (c) M3 and (d) M4.*

refractive index $n = 3.2$. The values are also found in Tab. 7.2 and we observe a good agreement in the frequencies. The Q factors however are many orders of magnitude larger than what we observe for the disk resonator. They are so large in fact, that our numerical root-finding algorithm can not resolve the vanishing imaginary part for some of the modes. Of course, the large discrepancy is readily explained by the presence of the waveguides, which offer an additional decay channel. Furthermore, the waveguides are also the key to understanding the increasing Q factor with increasing l. In this case, the mode is more localized in the center of the disk (see below) and simply has less overlap with the waveguide modes. For completeness, we also compare our resonances to the FDTD results published in Ref. [70]. We find that the frequencies agree very well with deviations below 1%. The Q factors on the other hand, are significantly reduced in the FDTD calculations. Particularly for the modes of higher radial order, the DG
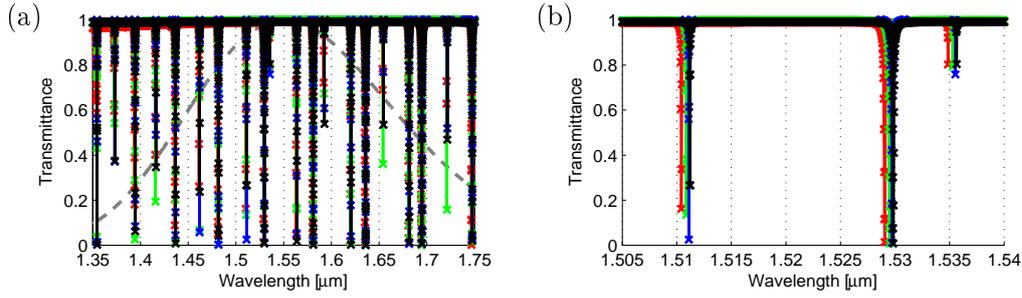
Figure 7.8: *Spectra for meshes M1 (red), M2 (green), M3 (blue) and M4 (black). All calculations were done in 4th order. The dashed line in (a) indicates the spectrum of the exciting pulse. Panel (b) shows a close-up of the spectrum.*
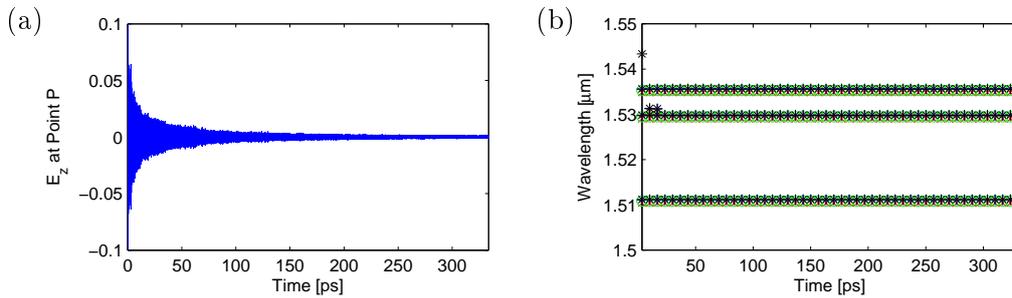
Figure 7.9: *The result of the filter diagonalization as a function of the runtime of the simulation. Different meshes are indicated by different symbols and colors: M1 ('x', red), M2 ('o', green), M3 ('+', blue) and M4 ('*', black).*

results are up to a factor 25 higher than those obtained via FDTD. This significant reduction of the Q factors for the FDTD calculations can only be attributed to the artificial surface roughness due to the staircasing, which introduces an unphysical Q_{surf}.

Finally, we study some of the modes in more detail, by exciting them individually. To this end, we replace the short waveguide pulse by a continuous-wave source. The frequency is tuned to the individual resonance and a snapshot of the field distribution is recorded after $T \approx 500$ ps. The field plots corresponding to the three resonances of Fig. 7.9(b) are depicted in Fig. 7.10. In addition, we also show the field for a case, where the excitation wavelength does not coincide with a resonance of the disk. From the field plots, we can directly recognize the azimuthal and radial order of each mode. Furthermore, we observe that the radiation is coupled from the upper waveguide into the disk and then leaves the system through the left port of the lower waveguide. While this effect is clearly visible in Fig. 7.10(a), it is less pronounced for the modes of higher radial order. This can be attributed to the high Q factors (cf. Tab. 7.2), which lead to extremely narrow resonances. Despite specifying the incoming wavelength to 6 digits, we still do not hit the resonance perfectly. The mode pictures are thereby consistent with the transmittance spectra in Fig. 7.8(b), where we only find a partial reduction
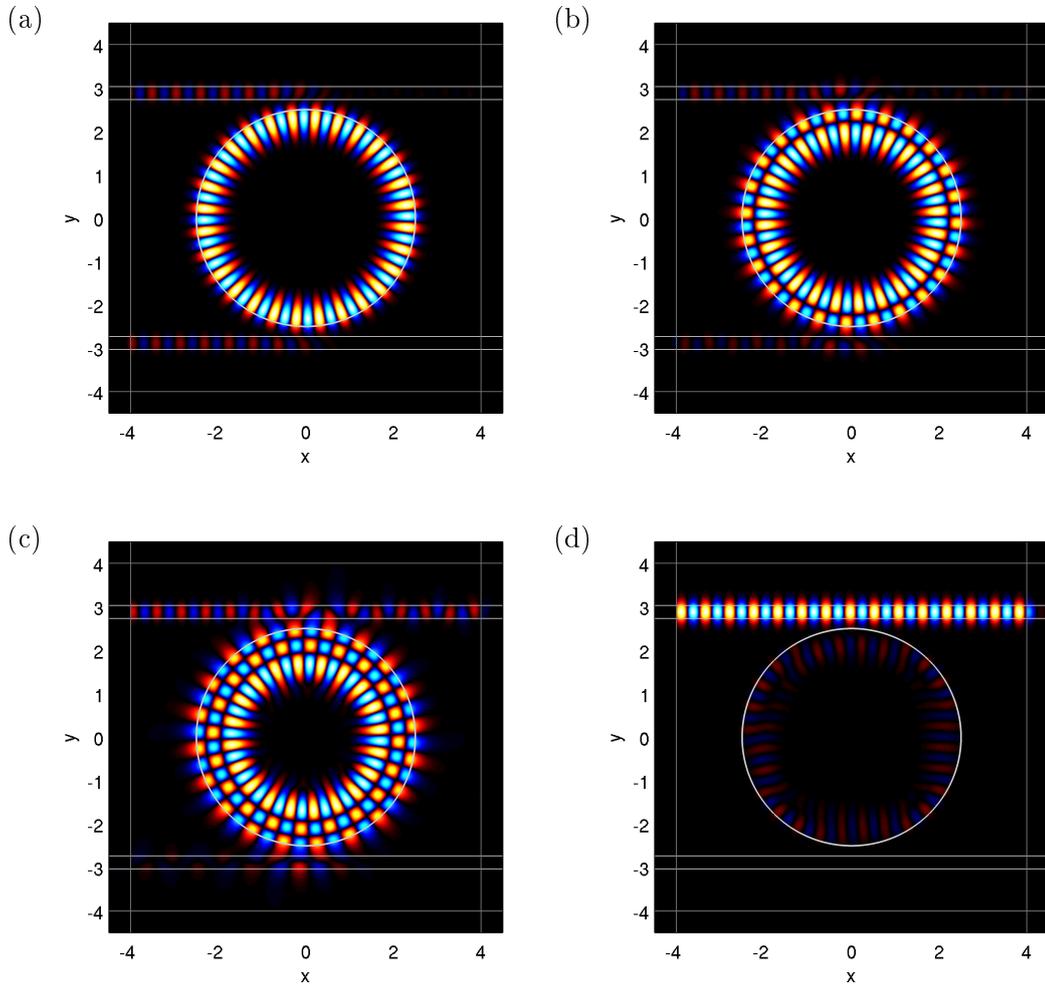
Figure 7.10: *Field plots of the E_z-component after excitation with a sinusoidal waveguide mode with different frequencies: (a) First-order radial mode ($m = 28$, $\lambda = 1.529\,\mu m$), (b) second-order radial mode ($m = 24$, $\lambda = 1.5112\,\mu m$), (c) third-order radial mode ($m = 20$, $\lambda = 1.5357\,\mu m$) and (d) off-resonant excitation ($\lambda = 1.5198\,\mu m$).*

m	l	DG		FDTD [70]		Mie's Theory	
		$\lambda\,[\mu\text{m}]$	Q	$\lambda\,[\mu\text{m}]$	Q	$\lambda\,[\mu\text{m}]$	Q
27	1	1.58113	9181.8	1.58451	7600	1.5817	2.13×10^{18}
28	1	1.52985	12286	1.53319	7500	1.5303	1.13×10^{19}
29	1	1.48185	16354	1.48521	7000	1.4823	$> 10^{19}$
30	1	1.43684	21865	1.44010	9100	1.4373	$> 10^{19}$
23	2	1.56388	28782.2	1.56724	8700	1.5644	5.78×10^{12}
24	2	1.51121	39050.1	1.51455	9900	1.5117	2.66×10^{13}
25	2	1.46210	52434.4	1.46543	11000	1.4625	1.23×10^{14}
19	3	1.59272	109861	1.59610	9900	1.5932	2.32×10^{8}
20	3	1.53565	89068.5	1.53901	9700	1.5360	9.02×10^{8}
21	3	1.48271	195808	1.48605	8800	1.4830	3.56×10^{9}

Table 7.2: *Numerical values for selected resonance frequencies of a disk resonator. The different colors indicate different calculations: DG (red), FDTD as taken from Tab. II in Ref. [70] (green) and a dielectric cylinder (with parameters $r = 2.5\mu m$, $n = 3.2$) according to Mie's theory (blue).*

in the transmittance for modes of higher radial order.

7.3.2 Ring Resonators

As we have demonstrated in the previous section, a disk resonator supports modes of higher radial order. While these modes can be useful for certain applications due to their higher Q factors, more often they pose a problem. Particularly, we can not realize a device with equidistant resonance frequencies, which is desirable for many filtering applications. Fortunately, there is a rather simple way to suppress the higher-order radial modes by etching out the center of the disk. If the resulting ring is sufficiently narrow, it will only support a single mode in radial direction. To study such a system in more detail, we model a ring as depicted in Fig. 7.11(a). In contrast to the previous structure, we now work with a two-port device. Our waveguide is assumed to be $w = 400\,\text{nm}$ wide and the outer radius is taken to be $r = 4\,\mu\text{m}$. Concerning the material, we use a refractive index of $n = 3$ and the whole system will be considered in TE polarization.

Since we are dealing with a two-port device, it makes no sense to look at the spectrum of the transmission through plane B. Eventually, all radiation will find its way out of the ring and, neglecting the very small losses, the transmittance will be 100% for all frequencies. Instead, we will concentrate on the dispersive properties of the system.

(a) (b)
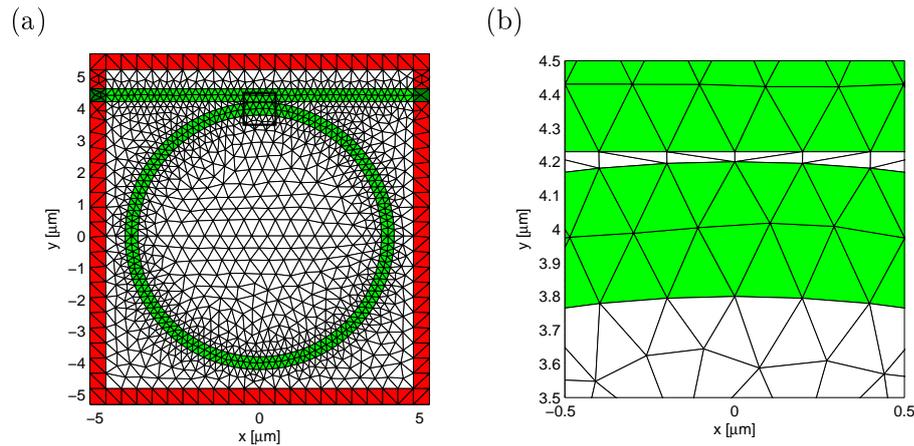
Figure 7.11: *(a) Mesh used to model a ring resonator with gap width $d = 50$ nm. (b) Close-up of the gap.*

Coupling Efficiency

One of the important characteristics of a resonator coupled to a waveguide is the power coupling coefficient K. It describes, what percentage of the energy will be transfered from the waveguide into the ring or vice versa. Intuitively, it is clear, that this coupling efficiency is strongly influenced by the distance between the ring and the waveguide. Furthermore, it depends on the radius and it is a function of the wavelength. To determine K numerically, we measure the flux across surface C (cf. Fig. 7.5(b)) and normalize it to the incoming flux measured at plane A. For this calculation, we need to take care that the simulation is terminated before the pulse has finished its first round-trip and returns to C for a second time. Therefore, sufficiently narrow pulses must be used for this study.

The power coupling coefficient K as a function of wavelength is plotted in Fig. 7.12. As expected, the coupling efficiency strongly increases if we reduce the distance between
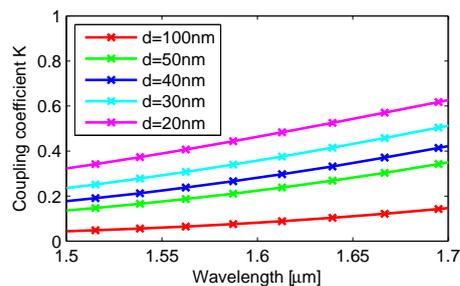
Figure 7.12: *Coupling efficiencies of ring resonators with varying distance d to the waveguide.*

ring and waveguide. As the gap between waveguide and ring gets smaller, the structure

becomes increasingly more difficult to resolve numerically. Here, the DG method is at its most impressive, since the adaptive mesh allows to locally refine the geometry (see Fig. 7.11(b)). All calculations were done with several refinements and expansion orders to ensure numerical convergence. The mesh shown in Fig. 7.11, in combination with a 4th order discretization was found sufficient to achieve results converged to at least three decimal places.

Group Delay

As discussed above, the all-pass setup does not allow for outcoupling of certain frequencies. However, the phase response of the system still is strongly frequency dependent, as the resonant parts of the spectrum will be delayed considerably. This introduces a so-called group delay, which allows us to use such devices for dispersion compensation [72, 73]. The group delay is also interesting with respect to all-optical buffering and delay lines [6].

Formally, the group delay is defined as

$$\tau(\omega) = -\frac{\partial \Phi(\omega)}{\partial \omega}, \tag{7.4}$$

where $\Phi(\omega)$ represents the phase difference between the left and the right port. For a lossless ring and in the limit $K \ll 1$, we can express this phase difference analytically as [73]

$$\Phi(\omega) = \tan^{-1}\left[\frac{K(\omega)\sin(\theta(\omega))}{2\sqrt{1-K(\omega)} - \big(2 - K(\omega)\big)\cos(\theta(\omega))}\right] \tag{7.5}$$

with the round-trip phase shift

$$\theta(\omega) = n_{\text{eff}}(\omega)\,\omega\,2\pi r. \tag{7.6}$$

Here, n_{eff} is the effective index of the waveguide, while r denotes the outer radius of the ring.

Thus, we have two options for calculating the group delay. First, we can numerically record the phase delay between the two ends of the waveguide. From a numerical differentiation according to Eq. (7.4), we can directly obtain the group delay. As a second option, we can exploit expression (7.5) to analytically calculate the group delay for a given coupling strength K. While the coupling must still be obtained numerically, the second method eliminates all influence of numerical dispersion and artificial surface roughness. In Fig. 7.13, we present the results obtained via both approaches. We find that for the DG results presented in Fig. 7.13(a), both approaches agree very well. The minor deviations can be attributed to the approximations used in deriving Eq. (7.5).

For comparison, we present results from the analysis with a commercially available FDTD package [74]. The data for two different grid sizes, $\Delta = 50\,\text{nm}$ and $\Delta = 25\,\text{nm}$, is shown in Fig. 7.13(b). Interestingly, we already find a deviation between the semi-analytic solutions of DG and FDTD. This discrepancy can be traced back to a difference

(a)

(b)
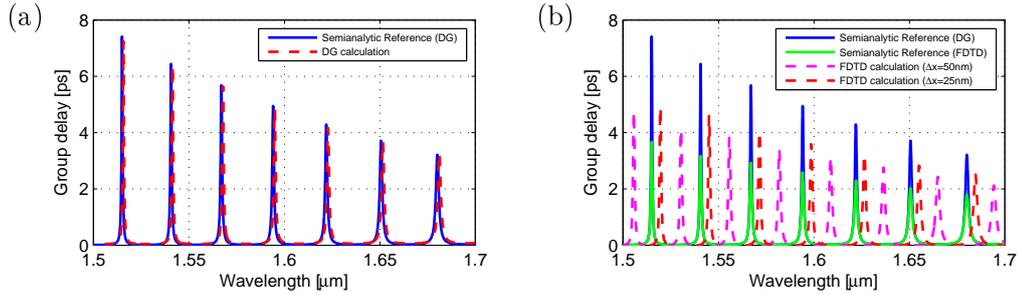
Figure 7.13: *Group delay for a ring of radius $r = 4\,\mu m$ and with a gap of $d = 50\,nm$. (a) Contains the results for a calculation with the DG method, while (b) displays data obtained with FDTD. In both cases, the semianalytic result, calculated via Eq. (7.5), is represented by the solid line. The dashed lines signify results obtained by numerical differentiation of the recorded phase delay.*

in the coupling strength K. It is readily explained by considering the gap width between ring and waveguide. Here, we only have a distance of 50 nm, which corresponds to either one or two cells in the FDTD simulation. The DG calculations on the other hand employ an adaptive mesh, which properly resolves the gap, even for distances below 20 nm (see Fig. 7.11(b)). Fortunately, the coupling only influences the height of the semi-analytic peaks. The position is solely defined by the geometrical parameters of the ring.

To account for the difference between semi-analytic and numerical results, we have to consider two effects. The first contribution again comes from an insufficient resolution of the geometry. Due to the finite size of the discretization cells, we have an uncertainty in the precise value of the radius for the discretized ring. Indeed, by varying the radius in Eq. (7.6) of the order of the lattice spacing, we manage to superimpose the semi-analytic with the numeric solution. However, there is a second contribution, which stems from the anisotropy of the phase velocity, as discussed in Sec. 4.2.2. There, we have demonstrated that the phase velocity varies for different propagation directions. For a second-order FDTD calculation, as employed here, the propagation is slower along the Cartesian axes than along the diagonals. As a consequence, the wave in the waveguide (parallel to the axes) is delayed when compared to the wave propagating in the ring (average over all directions). This leads to an effective round-trip phase-shift and therefore slightly changes the resonance frequency.

In conclusion, we have demonstrated that the DG method is well suited for the simulation of dielectric resonator structures. There are two distinctive advantages over the traditional FDTD method, which help to achieve accurate and well-converged results. The first relevant point are the adaptive meshes, which allow to precisely model arbitrary geometries. A second advantage is the higher-order nature of the method, which suppresses the effects of numerical dispersion. In combination, these feature allow to investigate large-scale dielectric structures with reasonable computational effort.

8 Chapter 8
Metallic Nanostructures

As we have seen in the previous chapter, the DG method fares quite well on calculations of dielectric resonators. As a next step, we proceed to metallic structures, which hold great potential for future applications (see Ch. 1). Unfortunately, as we will demonstrate, they also pose a significant challenge for some numerical methods.

In this chapter, we will discuss the numerical treatment of metallic structures in the time-domain. The fundamentals of how to incorporate dispersive media into a time-domain formulation were already covered in Ch. 3. Here, we will consider a few examples of realistic structures and apply our methods to study their optical properties.

We start this chapter with a brief theoretical introduction to plasmons. Next, we study the scattering properties of a metallic sphere, for which exact analytic results are available. After comparing the FDTD and the DG method, we proceed to an investigation of the transmission through sub-wavelength metallic slit-apertures. Finally, we consider two types of individual metallic nanoparticles, namely metallic rods and V-shaped structures. There, we are particularly interested in the field enhancements in front of the tip of such an object.

8.1 Plasmons and Surface Plasmon Polaritons

The relevant property of metals with respect to their optical response is the large free electron density. Very generally speaking, one can excite coherent oscillations in those free electrons, leading to charge density waves. These excitations are called plasmons and they can be launched in multiple ways. To us, only the excitation via electromagnetic radiation is relevant.

8.1.1 SPP Dispersion Relation

A specific kind of plasmons are the so-called surface plasmon polaritons (SPPs). They are hybrid oscillations of the electromagnetic field and the surface charge density at an interface between a dielectric and a metallic halfspace. The general setup is sketched in Fig. 8.1. From Eq. (2.17) we know, that the solutions of each halfspace are plane-
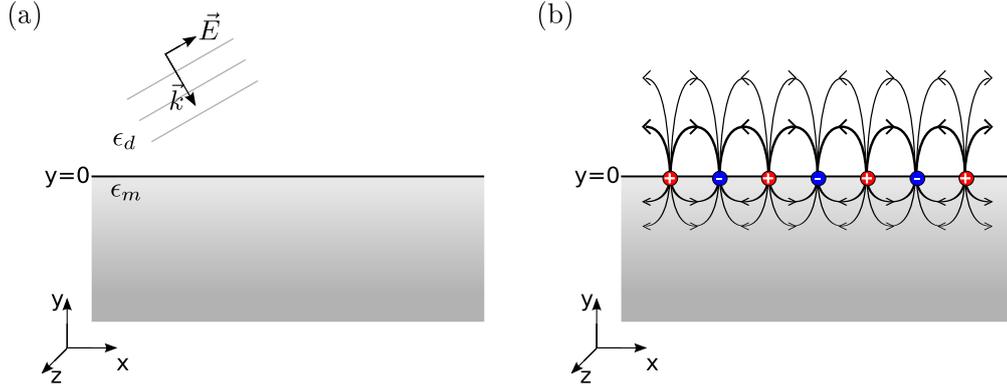
Figure 8.1: *(a) Sketch of a plane wave impinging on a flat metallic surface. (b) Schematic picture of the charge distribution and the corresponding electric fields at a metallic surface after excitation.*

waves, which need to be matched according to Eqs. (2.12). If we further require the waves to be localized at the interface, we can make the ansatz

$$\vec{E}_i(\vec{r}, t) = \vec{E}_0 e^{-\kappa_i |y|} e^{i(kx-\omega t)}, \tag{8.1a}$$

$$\vec{H}_i(\vec{r}, t) = \vec{H}_0 e^{-\kappa_i |y|} e^{i(kx-\omega t)}, \tag{8.1b}$$

where the index $i = m, d$ denotes the metallic or dielectric halfspace, respectively. Here, we have selected the propagation direction along the x-axis. Insertion of Eqs. (8.1) into Maxwell's equations then yields the dispersion relations

$$k(\omega) = \frac{\omega}{c} \sqrt{\frac{\epsilon_d(\omega)\epsilon_m(\omega)}{\epsilon_d(\omega) + \epsilon_m(\omega)}}$$

and

$$\kappa_i(\omega) = \sqrt{k^2(\omega) - \epsilon_i(\omega)\frac{\omega^2}{c^2}} = \frac{\omega}{c}\sqrt{\frac{-\epsilon_i^2(\omega)}{\epsilon_d(\omega) + \epsilon_m(\omega)}}.$$

Since we require the solution to be localized at $y = 0$ but propagating along the x-axis, we find the conditions

$$\mathrm{Im}\left[\kappa_i\right] = 0 \quad \text{and} \quad \mathrm{Im}\left[k\right] = 0,$$

which can be reformulated as

$$\epsilon_d(\omega)\epsilon_m(\omega) < 0 \quad \text{and} \quad \epsilon_d(\omega) + \epsilon_m(\omega) < 0.$$

Therefore, the two permittivities must have opposite sign and the modulus of the negative ϵ must be larger than the modulus of the positive one. As can be seen from
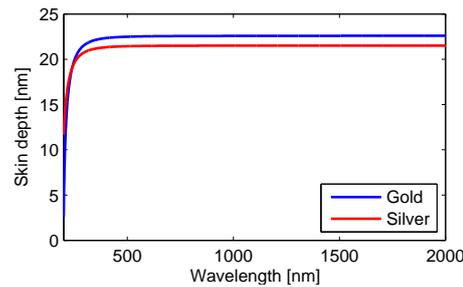
Figure 8.2: *The Skin depth of gold and silver as a function of the wavelength. For both cases, we assumed a Drude model with parameters taken from Tab. 3.1.*

the dispersion relation of gold and silver (see Fig. 3.2 in Sec. 3.1.2), these metals in combination with a standard dielectric fulfill above conditions in the red and near infrared spectrum. For silver, the conditions are even fulfilled for the entire visible range.

Since SPPs are localized at the metal-dielectric interface, we can ask for the penetration depth, i.e., the length scale over which the electric field decays inside the metal. Said length-scale is called Skin depth and it is given by $l = \mathrm{Re}\left[\frac{1}{\kappa_m}\right]$. Assuming a vacuum-metal interface and describing the metal via the Drude model allows to explicitly calculate l. In Fig. 8.2, we plotted the Skin depth as a function of the wavelength. From this plot, we recognize that l is almost constant in the visible and near-infrared spectrum. Furthermore, its exact value seems to be very similar for both materials and is roughly given by $l \approx 20\,\mathrm{nm}$. Thus, for devices of the lengthscale of the Skin depth, the electromagnetic field fully penetrates the structure.

8.2 Metallic Spheres

Before we can continue to study complex metallic nanostructures, we need to ensure that our numerical methods are actually capable of dealing with such objects. Similarly to the dielectric case, we employ a test system where the solution is known analytically. Here, it will be a metallic sphere of radius $r = 50\,\mathrm{nm}$. The material is assumed to be silver, which is implemented via a Drude model as described in Sec. 3.1.2.

Deviating from the dielectric case, we will use the scattering cross section C_{scat} of the metallic sphere for comparison. Similarly to the resonance frequencies, a reference solution $C_{\mathrm{scat}}^{\mathrm{ref}}$ can be obtained analytically via Mie's theory [67–69].

Calculations with FDTD

The setup for a calculation of the scattering cross section consists of a number of different regions as sketched in Fig. 8.3. In the center, we place a metallic sphere. Around this scatterer, we position a closed contour across which the electromagnetic flux is recorded. From the surface integral over the flux, we can later extract the
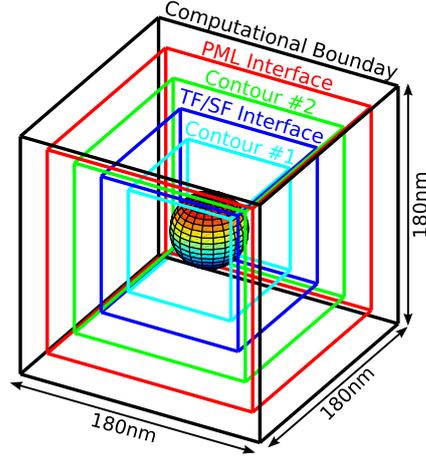
Figure 8.3: *Illustration of the scattering setup for the FDTD calculations.*

absorption cross section C_{abs}. Outside of this contour, we locate the TF/SF interface (see Sec. 4.3.2), which is used to inject radiation. Even further out, in the scattered field region, we place a second integration contour. The integrated flux through this surface results in the desired scattering cross section C_{scat}. To save computational time, this contour is placed at a distance of only 25 nm from the surface of the sphere. A far-field projection [28] is then employed to ensure an accurate calculation of C_{scat}. On the outside, we terminate the computational domain via PMLs. For the FDTD simulations, the whole system is discretized by an equidistant cubic Yee grid with spacing $\Delta := \Delta x = \Delta y = \Delta z$. The scatterer is illuminated by a broadband pulse with a carrier wavelength of $\lambda_0 = 200\,\text{nm}$ and the simulations run for $T \approx 1.67\,\text{ps}$. To get an impression on the required spatial discretization, we run calculations for a number of different cell sizes, ranging from $\Delta = 1.5\,\text{nm}$ down to $\Delta = 0.667\,\text{nm}$. The resulting cross sections are plotted in Fig. 8.4. It is immediatly evident that the agreement between the FDTD results and the analytical reference solution is very poor. Instead of the three pronounced resonances expected from Mie's theory, we find a multitude of spurious peaks and oscillations. To quantify the deviations, we introduce a measure for the relative error as a function of wavelength as

$$\mathcal{E}(\lambda) = \frac{\left| C_{\text{scat}}(\lambda) - C_{\text{scat}}^{\text{ref}}(\lambda) \right|}{\max_{\lambda}\left(C_{\text{scat}}^{\text{ref}}(\lambda) \right)}. \tag{8.2}$$

In Fig. 8.4(c), this error is plotted as a function of the wavelength. In the long-wavelength limit we do observe convergence, albeit reduced to first order due to the material interface (cf. Sec. 4.3.1). Closer to the resonances however, this convergence behavior disappears and we do not find a clear reduction of the error upon improving the discretization.

This effect can be understood by remembering the staircase approximation associ-
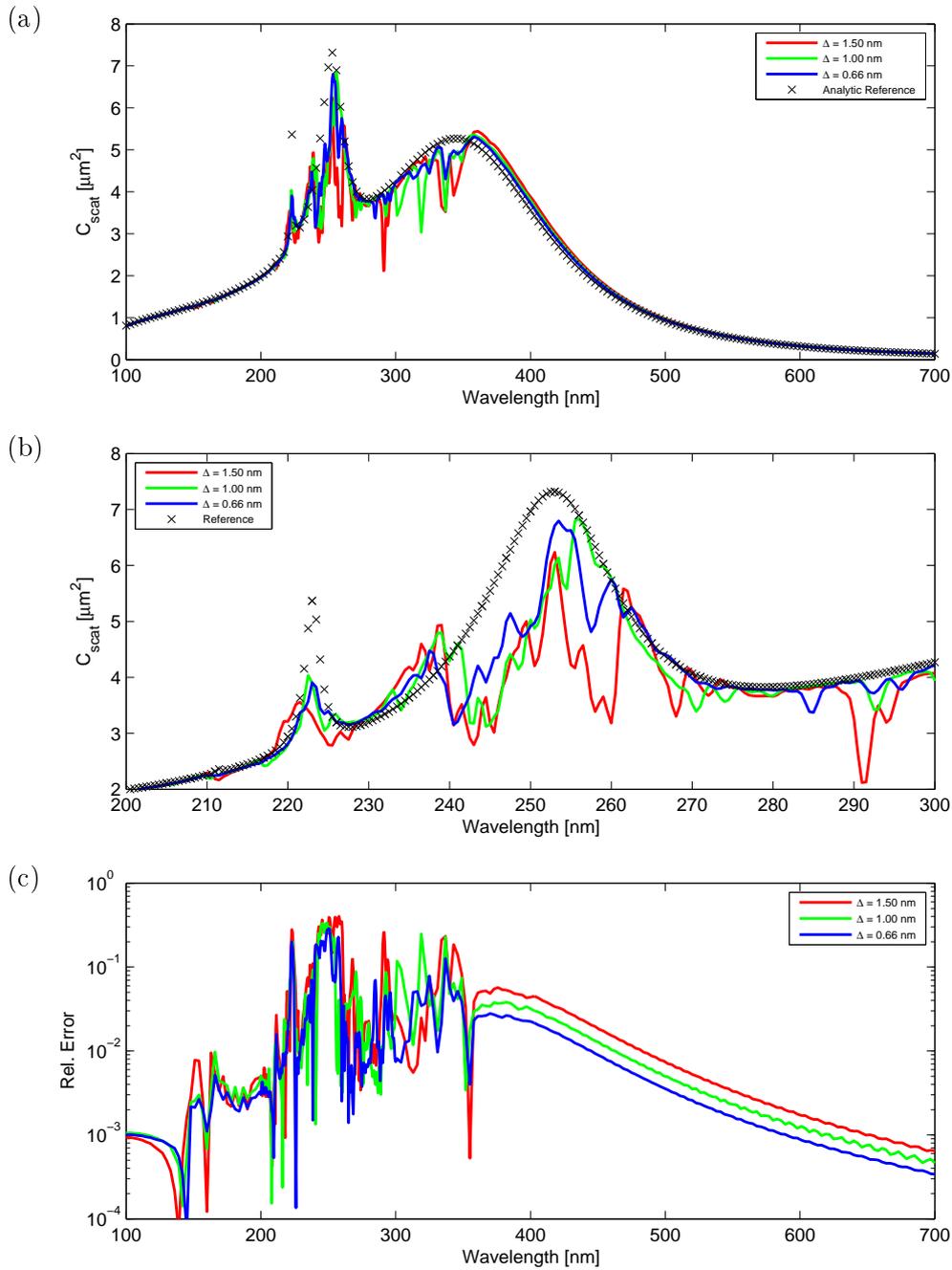
Figure 8.4: *(a) Scattering cross section of a silver nanosphere with radius r = 50 nm. (b) A close-up of the range between 200 nm and 300 nm. (c) A plot of the relative error as a function of the wavelength.*
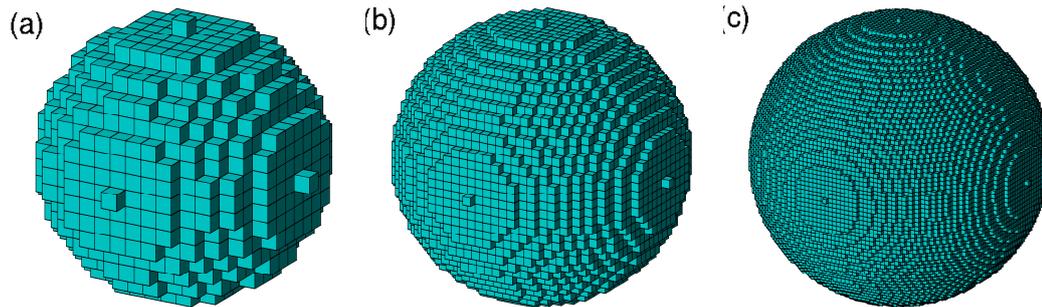
Figure 8.5: *Illustration of a discretized sphere for different refinements: (a) 8 cells per radius, (b) 16 cells per radius, and (c) 32 cells per radius.*

ated with the orthogonal mesh. Thus, the discretized sphere's surface actually consists of tiny metallic cubes, as illustrated in Fig. 8.5. Each of the small cubes has sharp corners, which theoretically lead to a singularity of the electric field at the tip [20]. Since the convergence analysis of the FDTD is based on the assumption of smooth fields, it breaks down in the vicinity of the singularities. Refining the grid does not necessarily improve the situation. On the contrary, it only introduces more (albeit narrower) singularities. We can confirm this explanation by looking at the fields obtained via an FDTD calculation. In Fig. 8.6 we plot the electric field component perpendicular to a cut plane through the center of the sphere. The plane is chosen such that the recorded electric field component is parallel to the polarization of the exciting wave. In the left column of Fig. 8.6, the analytical field obtained via Mie's theory is shown for three different wavelength. Next to it, in the right column, we present the corresponding field obtained from a FDTD calculation. The strong spurious peaks at the surface of the sphere are clearly visible. As a consequence, we have to conclude that the FDTD method is not well suited to calculate metallic nanostructures, if the geometry leads to an unphysical staircasing.

Calculations with the DG Method

To be able to compare the FDTD results with DG calculations, we generate a number of different meshes for a scattering setup similar to the one shown in Fig. 8.3. Since the surface of the sphere is expected to play a crucial role in the convergence, we employ meshes which are locally refined at the interface between sphere and vacuum. A cut through different discretizations is depicted in Fig. 8.7. In total we run nine different calculations, where we start from three basic meshes with edge lengths of $h_0 = 20\,\text{nm}$, $h_0 = 15\,\text{nm}$ and $h_0 = 10\,\text{nm}$ inside the sphere. Each of these three meshes is refined at the surface of a sphere by a factor $\alpha = 1, 2, 3$. The order of the expansion is chosen to be $N = 3$. The rest of the setup is very similar to the one of the FDTD calculations, with the difference that we do not require specific surfaces for the flux integration. Instead, we reuse the inner and outer side of the interface between the total field and

(a)

(b)

(c)

(d)

(e)

(f)
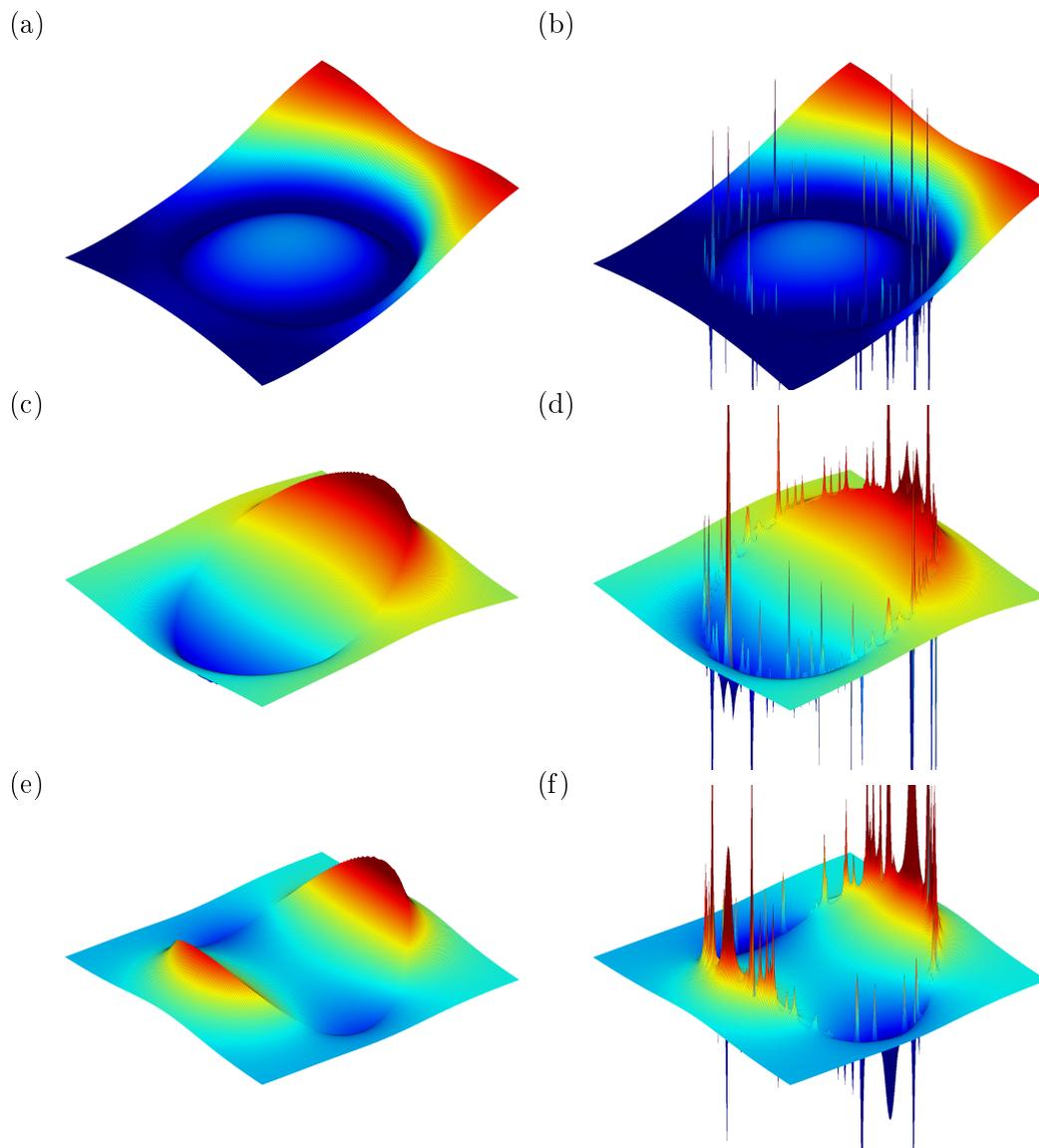
Figure 8.6: *Electrical field component perpendicular to the cutplane. The plots on the left represent the analytical results, while the plots on the right contain the corresponding FDTD data. The analysis was conducted for three different wavelengths: (a),(b) $\lambda = 346\,\text{nm}$, (c),(d) $\lambda = 253\,\text{nm}$ and (e),(f) $\lambda = 223\,\text{nm}$.*

(a)   (b)   (c)
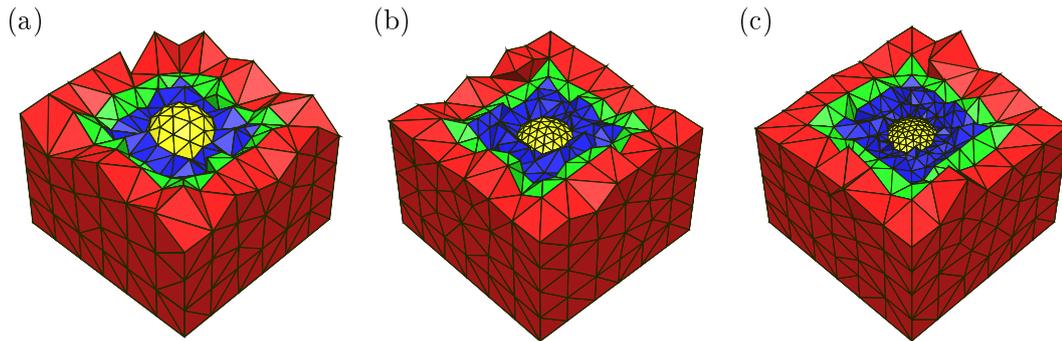
Figure 8.7: *Meshes for calculating the scattering cross section of a metallic nanosphere. Depicted is the coarsest mesh with $h_0 = 25\,\mathrm{nm}$ for three different refinements: (a) $\alpha = 1$, (b) $\alpha = 2$, (c) $\alpha = 3$. Different colors represent the different regions: Scatterer (yellow), total field region (blue), scattered field region (green), PMLs (red).*

the scattered field region.

The results of our calculations can be found in Fig. 8.8(a). At first sight, all calculations seem to reproduce the analytic result rather well. In particular, all curves are smooth and no spurious peaks are introduced. However, if we take a closer look at the resonance at $\lambda = 223\,\mathrm{nm}$ (Fig. 8.8(b)), we find that all calculations struggle to reproduce the peak accurately. Still, we clearly see how a refinement of the sphere's surface brings our results closer to the analytical solution. The relative error is plotted in Fig. 8.8(c). As expected, the error has a distinct peak at $\lambda = 223\,\mathrm{nm}$, which corresponds to the very narrow resonance discussed above. This feature aside, we observe reasonable convergence and the average error is below 1%. The effect that the error does not continue to decrease with better resolution can be attributed to the PMLs. We have demonstrated in Sec. 5.5.3, that a PML of only one cell for a calculation with $N = 3$ leads to reflections of the order 5×10^{-3}, which is roughly the number we observe here. Clearly, the DG method does not live up to its full potential in this example. It seems that the approximation of the spherical surface, despite the local refinement, is still not sufficient to yield the optimal convergence behavior for the resonance around $\lambda = 223$ nm. In order to improve this situation, one has to implement curvilinear elements. Furthermore, one could reduce the overall error by adding more cells to the PMLs or by increasing the order. Nonetheless, the DG method fares very well when compared to the FDTD calculations. The fact that we observe no spurious peaks or oscillations is very reassuring and allows us to employ the method for further studies in the following sections.

8.3 Transmission through Metallic Nano-Apertures

For our first calculations of an experimentally relevant system, we consider a thin metallic film with holes of subwavelength size in it. This system has attracted a lot of

(a)

(b)

(c)
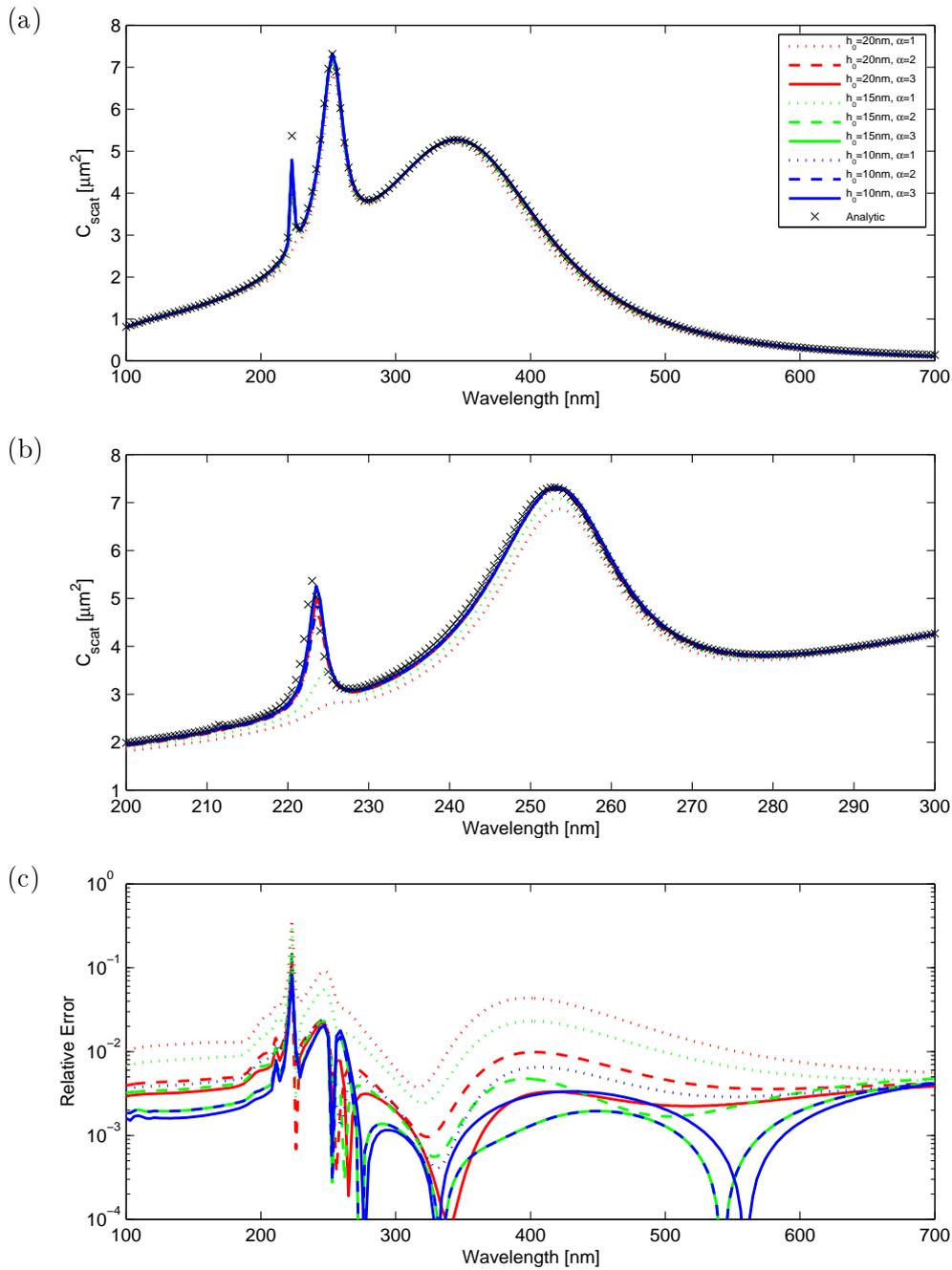
Figure 8.8: *Results of the DG calculations for a metallic nanosphere with radius r = 50 nm. (a) Scattering cross section over a broad spectral range. (b) A close-up of the region between 200 nm and 300 nm. (c) A plot of the relative error as a function of the wavelength. The colors belong to different edge lengths h_0, while the linestyles indicate the refinement. The black crosses denote the analytical reference solution.*
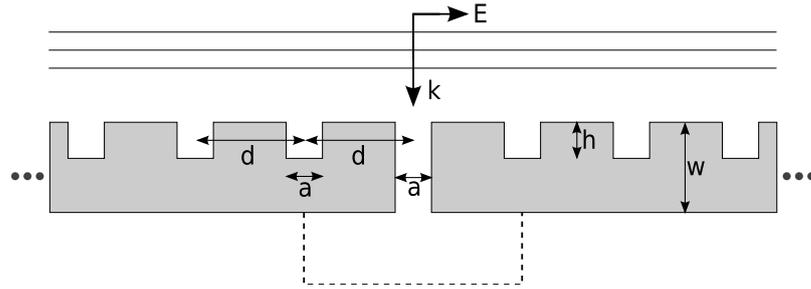
Figure 8.9: *Schematic setup of a slit aperture in a metallic film. The aperture is flanked by a finite number of grooves and the dashed line indicates a contour used to calculate the transmittance.*

attention in the past years, because it may exhibit extraordinary optical transmission [12]. From the start, it was suspected that the excitation and propagation of SPPs must play a major role in the explanation of the effect. This assumption is supported by a number of theoretical studies (e.g. [75–77]). Unfortunately, most theoretical work on extraordinary optical transmission is still based on approximations that describe the metal as a perfect electric conductor (PEC) [75, 76]. Although this simplifies the calculations and may provide a qualitative picture of the physical processes, it is not suitable for designing and optimizing experiments. As an application of the DG approach we will, in the following, study the influence of a more accurate description of the material on the transmittance properties of such apertures. A typical structure is sketched in Fig. 8.9 and consists of an infinite metallic screen of thickness w with a nano-structured surface. We assume a slit of width a going through the metal and next to it N grooves of width a and depth h that are spaced a distance d apart from each other. The many parameters make it obvious, that a fast and reliable numerical method is required to thoroughly study the properties of such systems. Although these structures can be grid-aligned and, thus be treated with FDTD, the DG method still exhibits a distinct advantage: The unstructured meshes allow an accurate resolution of the small nano-structures without wasting grid points in the bulk regions. Specifically, we employ a 4th-order discretization in all forthcoming calculations. Before getting to the full structure we first analyze the simple slit aperture without grooves so as to obtain a reference. We assume a metallic film of thickness $w = 350$nm with a single slit of width a and irradiate it with a plane wave, where the magnetic field is parallel to the slit. The pulse shape is taken to be a broadband Gaussian with a carrier wavelength of $\lambda_0 = 500$nm that covers the relevant spectrum from 350nm to 1000nm. Behind the aperture, we integrate the flux over the entire half-space and normalize it to the width of the aperture. First, we conduct a series of calculations with different slit width a for both, a PEC structure as well as for a more realistic Drude model with silver parameters $\omega_{pl} = 1.37 \times 10^{16}$s^{-1} and $\omega_{col} = 8.5 \times 10^{13}$s^{-1} [78]. We display the results in Fig. 8.10. Already for the simple slit-aperture, the differences between PEC and a Drude metal are quite significant. For the realistic
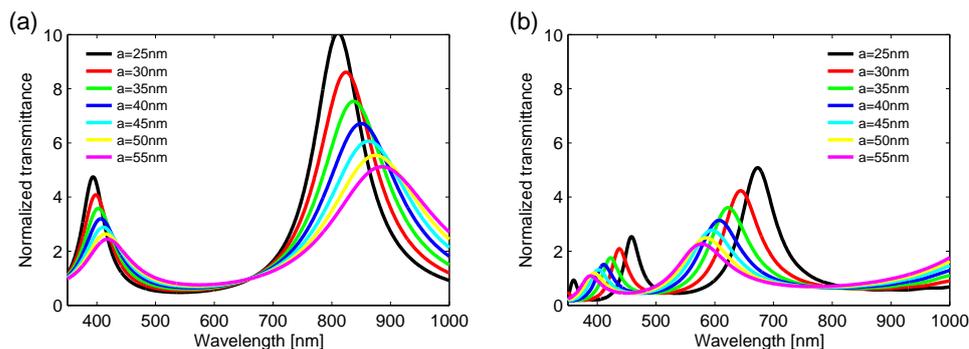
Figure 8.10: *Normalized transmittance spectra for simple apertures with differ-ent slit widths a. In panel (a) the spectra for systems with perfectly conducting metal are shown, while panel (b) contains the results for metallic systems that are described by a Drude model (see text for details).*

metal, the transmission is reduced, which is consistent with the fact that the silver exhibits absorption through an imaginary part in the dielectric function. Further, the position of the resonance at higher wavelength is strongly blue-shifted for the Drude case. Finally, for both material models the resonances shift with varying slit width a, but they move in opposite directions. For the PEC structure, the resonances shift to longer wavelengths as a increases, while they move towards shorter wavelengths when the Drude model is used to describe the metal.

We now proceed to the full structures with added grooves, and again start the analysis by comparing the PEC and the Drude model. The geometrical parameters are fixed to $d = 500$nm, $h = 100$nm, $w = 350$nm, and $a = 40$nm, identical to those reported in Ref. [76]. As discussed above, a simple slit shows two pronounced resonances, while adding the corrugation gives rise to a third peak with significantly increased transmittance. Our data for the PEC structure is in perfect agreement with results obtained via the modal expansion technique (cf. Fig. 2(a) in [76]). By comparing the data in Fig. 8.11(a) and 8.11(b), we again observe the strong influence of the material model on the spectra. In addition to the resonance shifts observed before, for the PEC case, the central resonance quickly saturates for $N \approx 5$ grooves. When using the Drude model, around $N = 15$ grooves are required to observe a similar enhancement. In a next step, we want to study the influence of the groove depth h and the spacing d between the grooves on the systems' transmittance properties. In accordance with the experimental setup in Ref. [76], we fix the number of grooves to $N = 5$ and keep all other parameters identical to those of the previous calculations expect for the height h, which is now scanned from $h = 20$nm to $h = 80$nm. All calculations are carried out with the more realistic Drude model. From the results depicted in Fig. 8.12(a), we deduce that changing the height results in a shift of the central resonance. The highest transmittance is observed, when this central resonance coincides with a peak of the simple aperture at 610nm (see Fig. 8.10(b)). The results for scanning the distance d between the grooves display a similar behavior as shown

(a)

(b)
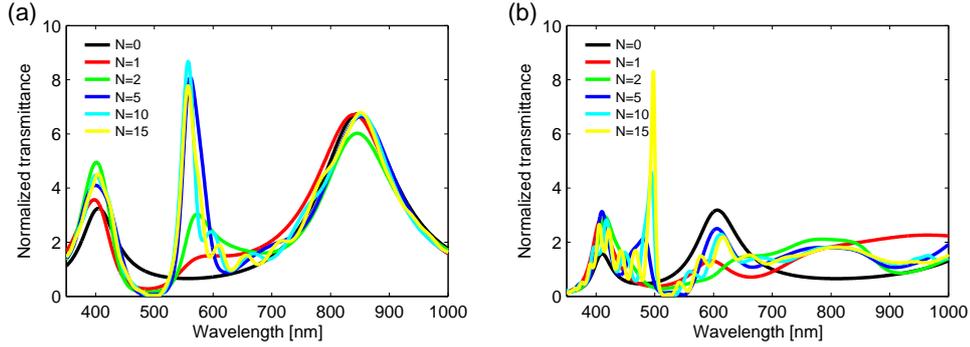
Figure 8.11: *Normalized transmittance spectra for slit apertures with added grooves for different numbers of grooves. The geometrical parameters are $d = 500nm$, $h = 100nm$, $w = 350nm$, and $a = 40nm$. Panel (a) depicts the spectra for systems with perfectly conducting metal, while in panel (b) the metal is described by a more realistic Drude model.*

(a)

(b)
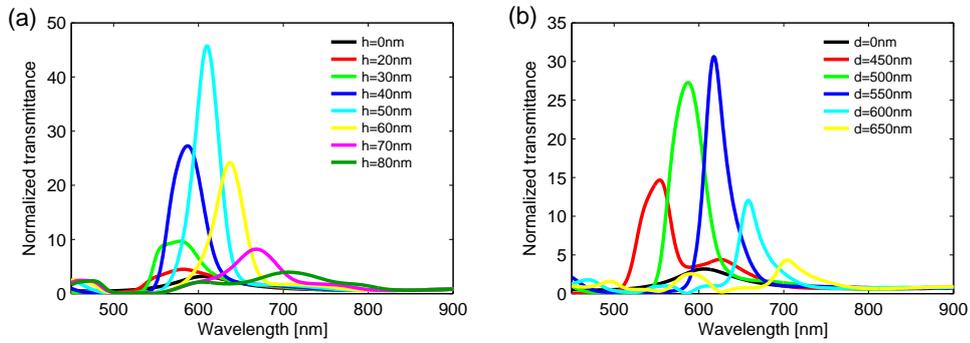
Figure 8.12: *Normalized transmittance spectra for different parameters of the corrugations. In (a), we shows the variation with the groove depth h, while (b) displays the results of changing the distance d between the grooves.*

in Fig. 8.12(b). Here, we have fixed the depth to $h = 40nm$ and changing d again leads to shifts of the central resonance frequency. As before, one observes that the maximum transmittance is reached when the resonance coincides with the resonance of the simple aperture.

8.4 Metallic Rods and V-Shaped Particles

Finally, we proceed to the analysis of three-dimensional metallic nanoparticles. Here, we focus on two particular geometries, both of which have been proposed in the context of local field enhancement. As discussed in Sec. 8.2, one has to be very careful when studying fields in the direct vicinity of metallic surfaces, especially if the surfaces are curved. In order to avoid problems from staircasing, we pursue all of the following
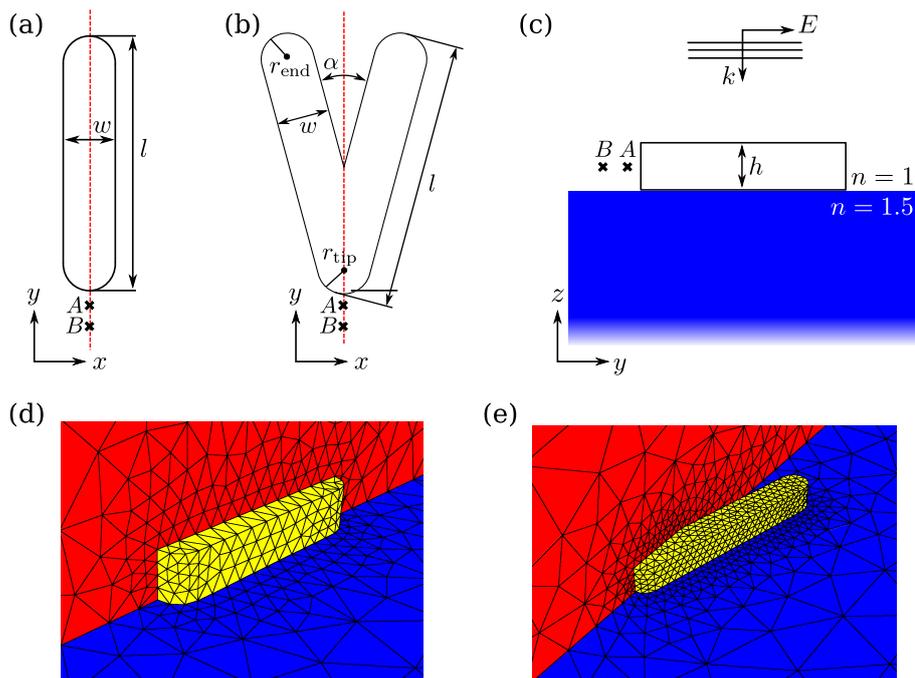
Figure 8.13: *Sketch of the (a) nanorod and (b) V-structure under consideration. Panel (c) depicts a sideview on the general setup with a plane wave impinging from the top. The metallic structures are assumed to rest on a glass substrate. In (d) and (e), we show the meshes used to model the nanorod and the V-structure, respectively. The red-shaded triangles in (d) and (e) represent the symmetry plane exploited to accelerate the numerical computation.*

calculations with the DG method.

8.4.1 The Nanobar

As a first system, we consider a silver nanobar as sketched in Fig. 8.13(a). This particular structure has the advantage that it has only few free geometrical parameters and can be easily fabricated [79]. To keep our calculations realistic, we assume that the bar is situated on a glass substrate. Furthermore, we chamfered the corners of the structure in order to avoid sharp metallic tips. This is in accordance with experimental realizations, where the corners of the rods always appear rounded (cf. electron micrographs in Ref. [79]).

Since we will only excite the structure under normal incidence and with the electric field polarized along the rod, we can exploit a mirror symmetry in order to accelerate the numerical calculations. The corresponding mirror plane is indicated in Fig. 8.13(d) and is implemented by enforcing a PMC boundary condition. The excitation itself consists of a short Gaussian pulse of approximately 10 fs duration. The corresponding
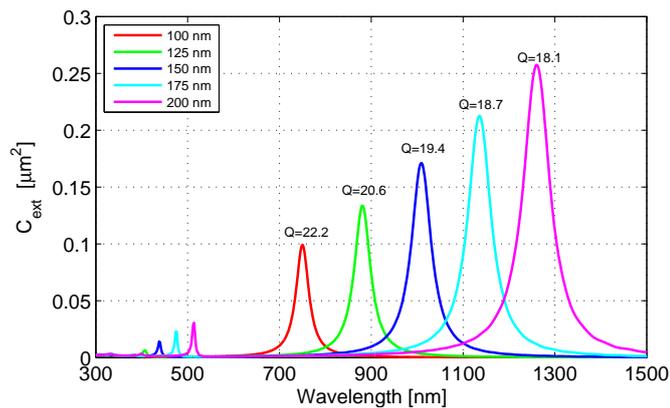
Figure 8.14: *Extinction cross section of nanobars of different lengths l. Each fundamental resonance is annotated with its corresponding Q factor. The small peaks on the left can be attributed to higher-order modes of the structures.*

spectrum ranges from $300\,\mathrm{nm}$ to $1500\,\mathrm{nm}$. After the excitation, we record the fields at the interface between the total-field and the scattered-field region. Due to the discontinuous nature of our method, we can record both, the fields inside and outside of the TF/SF contour. From the field data, after a Fourier transform, we obtain the Poynting vectors in frequency domain. Integration over the entire surface then yields the total flux. Normalization with respect to the incoming spectrum results in the absorption cross section and the scattering cross section for the fields inside and outside of the contour, respectively. The sum of the two cross sections finally yields the extinction cross section C_{ext} which we use for the characterization of our nanoparticles.

In a first set of calculations, we fix the height and width of the rod at $h = w = 20\,\mathrm{nm}$ and vary the length l between $100\,\mathrm{nm}$ and $200\,\mathrm{nm}$. The chamfering of the corners was achieved by replacing the endfaces of the bars with half-cylinders. From the results depicted in Fig. 8.14, we find that the resonance shifts to shorter wavelength as we decrease the length of the rod. This was to be expected and in a very crude approximation, we can interpret the nanorod as a small antenna. However, in contradiction to classical antennas, the fundamental resonance wavelength is found to be much longer than twice the rod length. In order to extract the properties of the individual resonances, we fit a Lorentzian curve to each peak (cf. Sec. 7.1). The extracted resonance wavelengths and Q factors are plotted in Fig. 8.15. The quality factors are found to vary only weakly with the length and a saturation can be observed for very small lengths. This is in good agreement with theoretical expectations for the quasi-static limit [80].

With the knowledge of the resonance wavelengths, we now proceed to the second part of the analysis. There, we excite the structure with monochromatic light at its fundamental resonance. During the illumination, we record the electric field at two points A and B in front of the nanobar (see Fig. 8.13(a)). The recording sites are located at a
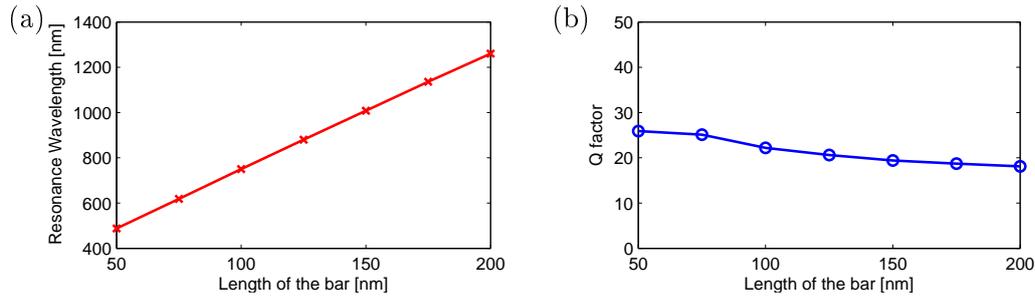
Figure 8.15: *(a) The resonance frequencies and (b) Q factors as a function of the length of a nanobar. Width and height of the structure are kept constant at $w = h = 20\,nm$.*
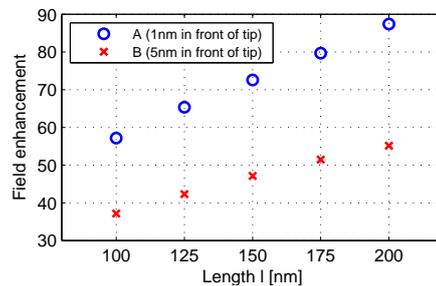
Figure 8.16: *Maximal field enhancement observed a point A (circles) and point B (crosses).*

distance of $1\,nm$ and $5\,nm$ from the tip of the structure, respectively. By studying the recorded signals, we find that the maximal field strength saturates after approximately 30 optical cycles. Thus, we store the maximal amplitudes after 40 cycles and plot them as a function of the length l in Fig. 8.16. Interestingly, the field enhancement seems to increase almost linearly with the length of the bar. However, we can not increase the enhancement arbitrarily, since the resonance wavelength also increases with the length of our structure. Thereby, we are restricted to field enhancements of approximately 100 if an excitation wavelength around $1500\,nm$ is desired. Furthermore, we need to keep in mind that the recording point A is only $1\,nm$ away from the structure and the enhancement decays exponentially with the distance. Still, a factor 50 in the field results in a factor 2500 for the intensity. For many applications such as surface enhanced Raman spectroscopy (SERS), this can already lead to a significant improvement of the sensitivity.

8.4.2 Metallic V Structure

In comparison to the nanobar, we now consider a V-shaped structure as sketched in Fig. 8.13(b). This shape was originally proposed by Mark Stockman and coworkers
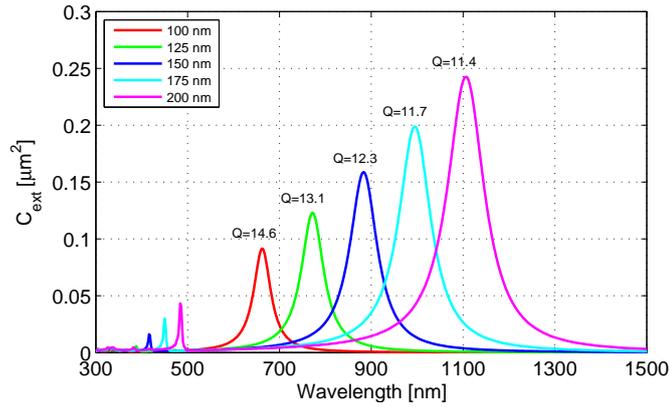
Figure 8.17: *Extinction cross section of V-shaped metallic nanostructures for different armlengths.*

[81] with the sole purpose of creating strong field enhancements at its tip. It should be noted that the analysis of Stockman et al. was conducted in the quasi-static limit, where all retardation effects are neglected. Furthermore, the authors of Refs. [81, 82] employ chirped excitation in order maximize the field enhancement.

Here, we will study a structure with experimentally realistic dimensions. Therefore, our structure is barely at the edge of the onset of the quasi-static regime [83] and we can not expect quantitative agreement with the results of Ref. [81]. Furthermore, we will refrain from using chirped pulses, to limit the number of free parameters.

Similarly to the nanobar studied previously, the width and height of the V-shape are fixed at $w = h = 20$ nm. The apex angle is chosen as $\alpha = 30°$ and we assume the structure to be located on a glass substrate. As before, the chamfering of the endfaces and the tip is achieved by replacing them with half-cylinders of radius $r_{\mathrm{rounding}} = w/2$. The excitation is performed identically to the one used for the nanobar, and the polarization is chosen to be parallel to the V's symmetry axis. Therefore, we can again exploit the mirror symmetry to reduce the computational effort by a factor of two.

In Fig. 8.17, we show the extinction cross sections for a number of V-shaped particles of different armlength l. By comparing the curves for the V-structure with those of the nanobar (Fig. 8.14), we find a very similar scaling behavior. Both structures show a diminishing extinction cross section for shorter lengths. As expected, the resonance wavelength also reduces with shrinking size. The only obvious difference in the spectra of particles with identical lengths consists in a slightly shorter resonance wavelength and in a lower Q factor for the V-structure (cf. Fig. 8.15 and Fig. 8.18).

In a second step, we proceed to evaluate the field enhancement in front of the rounded apex. The general procedure is identical to the case of the nanobars and the resulting data for the two recording points A and B is plotted in Fig. 8.19. The first thing to note is that the enhancement is significantly lower for a V structure than for a nanobar. Looking at the exact numbers, we find the enhancement of a nanobar to be between
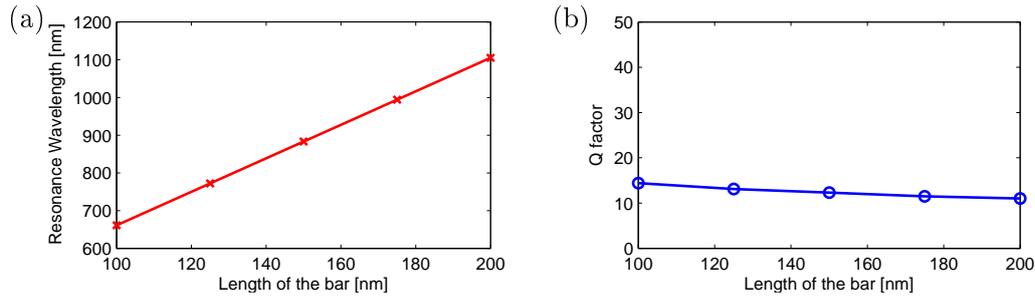
Figure 8.18: *(a) The resonance frequencies and (b) Q factors as a function of the length of a metallic V structure. Width and height of the structure are kept constant at $w = h = 20\,nm$. The apex angles is $\alpha = 30°$.*
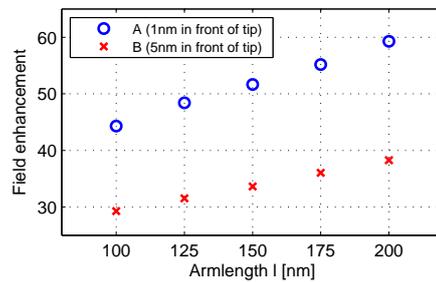
Figure 8.19: *The field enhancement in front of (a) the nanobar and (b) the V-structure. The different symbols correspond to different recording points (cf. Fig. 8.13).*

30 % and 50 % higher than that of the equivalent V-shape.

To visualize the spatial variation of the enhancement, we finally depict the maximal intensity in the vicinity of the V-structure in Fig. 8.20. The data depicted here was calculated for a V-shape with an armlength of $l = 150\,nm$. From this data, we find similar enhancement at the tip and at the endfaces of the V. Furthermore, we observe the exponential decay with distance from the surface (note the logarithmic colorscale). At 10 nm above the structure (Fig. 8.20(b)). Finally, we would like to note that more detailed studies of metallic nanorods and V-structures, conducted with our DG method, can be found in the masters thesis of Kai Stannigel [83]. There, he also investigated the influence of the various geometric parameters such as apex angle, width and height of the structures. Furthermore, a thorough investigation of the excitation via chirped pulses was conducted. For all the studied cases, no significant enhancement over the values reported here could be found.

inputModifiedRadiationDynamics/ModifiedRadiationDynamics.tex
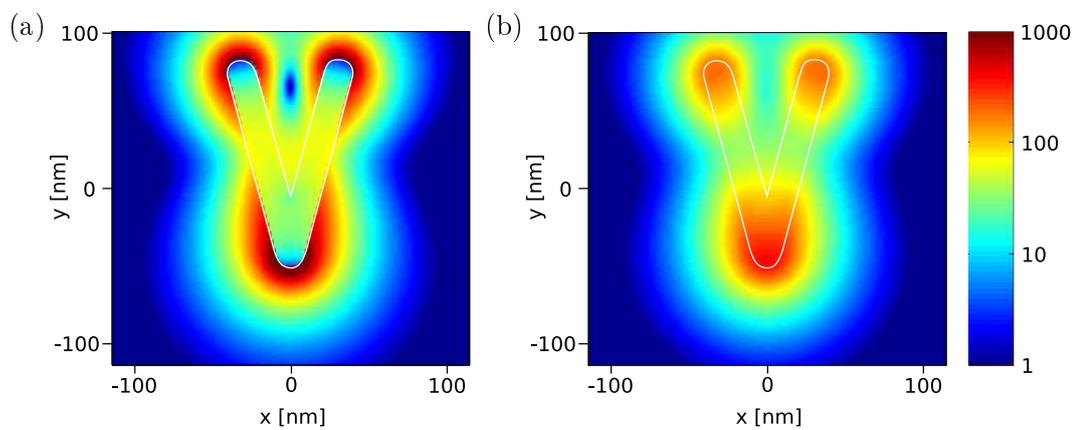
Figure 8.20: *Maps of the intensity enhancement in the vicinity of a V-shaped metallic nanostructure (indicated by the white line). Panel (a) shows the enhancement factor on a horizontal cut through the center of the V. In (b) we depict the factor in a plane 10 nm above the structure.*

9

Summary and Outlook

One of the key requirements for the design and optimization of nanophotonic devices is the availability of accurate, flexible, and efficient computational methods. In order to keep up with the steady progress in fabrication technology, it is certainly not sufficient to purely rely on Moore's law[1]. Instead, there is a pressing need for new and more powerful numerical approaches.

Summary

The aim of the present thesis was to investigate different ways to improve the time-domain simulations of nanophotonic devices. After a short introduction and motivation in chapter 1, we started with a brief recapitulation of Maxwell's equations and some of their properties (chapter 2). From the basic constitutive equations, it became obvious that a time-domain approach has the disadvantage that one can not directly include dispersive materials. This constitutes a severe limitation, since one of the most promising material classes in nanophotonics—metals—is strongly dispersive. To overcome this obstacle, we used chapter 3 to introduce the method of auxiliary differential equations (ADEs), which allows to treat dispersive behavior within a time-domain framework. Besides dispersion, this technique also facilitates the implementation of absorbing boundary conditions, thereby enabling us to model open systems.

With the theoretical framework set, in chapter 4 we discussed the workhorse of all ab-initio methods for Maxwell's equations, namely the finite-difference time-domain (FDTD) method. Besides the fundamentals of this approach, we also studied a possible extension towards higher-order discretizations. However, in a set of numerical experiments we demonstrated that this improvement only works for fully homogeneous systems. In chapter 5, we then introduced a more sophisticated spatial discretization, the discontinuous Galerkin (DG) scheme, which manages to overcome some of FDTD's inherent limitations. In particular, we showed that—in contrast to the FDTD

---

[1]Moore's law describes the empirical observation that the number of transistors in a typical computer processor doubles approximately every two years. It thereby indicates an exponential growth of the available computing power over time.

method—the DG scheme allows for higher-order approximations in the presence of material interfaces. This feature essentially allows us to obtain much more accurate results with less computational effort. For high-precision calculations, we have demonstrated that results can be obtained several orders of magnitude faster with the DG method when compared to FDTD.

In chapter 6, we deviated from the question of spatial discretization and turned to the problem of time integration. In particular, we studied a linear matrix-exponential integrator based on a Krylov-subspace approach. Besides the bare scheme, we also provided essential extensions to make the scheme suitable for practical applications. Finally, in a series of numerical experiments, we demonstrated the method's attractive features of strong stability and high accuracy.

With all the different numerical methods at hand, in chapters 7 and 8, we applied them to a number of real-world challenges. The first tests were conducted on two-dimensional dielectric resonator structures. A direct comparison of an FDTD calculation with DG results for an analytically solvable cylinder already hinted at the superiority of the DG method. This observation was then confirmed by studies of a disk and a ring resonator structure. To preclude any bias from our side, we conducted these comparisons using established results from the literature as well as calculations with a commercially available FDTD package.

For the final set of applications, in chapter 7, we then proceeded to the analysis of metallic nanostructures. Again, we started with an analytically solvable problem to allow for a quantitative study of the numerical errors. In this case, we calculated the scattering cross section of a small silver sphere. Here, the FDTD method fared much worse than for the dielectric resonators. It essentially produced an unusable spectrum, littered with spurious peaks and unphysical oscillations. The DG method in contrast had no problems with these particular calculations and yielded accurate results. The strange behavior of the FDTD method was explained as an artifact of the insufficient geometrical approximation. After this analytic comparison, we proceeded to realistic experimental setups. In particular, we investigated the transmission through a sub-wavelength aperture. Here, we demonstrated that for nanophotonic systems, the proper description of the materials is essential. Concretely, we showed that describing a metal as a perfect electric conductor can lead to qualitatively wrong results. In a final study, we investigated the capability of small metallic nano rods and V-structures to generate strong local field enhancements. Interestingly, we did not find particularly strong enhancements at the tip of V-shaped structures. Quite the contrary, a simple nanorod of similar extent can provide up to 40% stronger fields.

Conclusions and Outlook

In conclusion, we have investigated both, the influence of the spatial discretization as well as the effect of the time-integration on numerical simulations of photonic nanostructures. Out of these two factors, the spatial discretization was identified as the dominating cause of numerical errors. Therefore, we proposed a discontinuous Galerkin scheme as a flexible, accurate and fast alternative for the commonly applied FDTD

method. In a series of realistic calculations, this approach has proven itself successful. In view of its attractive properties, we expect the DG scheme to soon become a popular tool in the repertoire of numerical methods for nanophotonic systems.

In addition to its already competitive performance, the DG method also holds great promise for further improvements. Possibly the most relevant extension for nanophotonic calculations is the implementation of curvilinear elements. This feature would allow for an exact representation of rounded structures and should thereby further increase the accuracy and performance of the method. A second point where we can still expect significant improvements is the time-integration. While we established that the spatial discretization is responsible for most of the numerical errors, the performance is still strongly limited by the time-integration. We expect that an advanced combination of implicit and explicit methods could lead to a dramatic reduction of the computational time. Last but not least, we would also like to see more sophisticated material models implemented in the DG approach. This includes but is not limited to classical nonlinearities, anisotropic media or active materials such as quantum dots.

Bibliography

[1] R. Mausfeld. *Farben. Betrachtungen aus Philosophie und Neurowissenschaften*, chapter Zur Natur der Farbe. Die Organisationsweise von Farbe im Wahrnehmungssystem. Surkamp, 2007.

[2] J. C. Maxwell. A dynamical theory of the electromagnetic field. *Philosophical Transactions of the Royal Society of London*, 155:459–512, 1865.

[3] A. Einstein. über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt. *Ann. Phys.*, 322:132–148, 1905.

[4] D. E. Chang, A. S. Sorensen, E. A. Demler, and M. D. Lukin. A single-photon transistor using nano-scale surface plasmons. 2007. URL `arXiv:0706.4335v1`.

[5] W. Cai, U. K. Chettiar, A. V. Kildishev, and V. M. Shalaev. Optical cloaking with metamaterials. *Nature Photon.*, 1:224–227, 2007.

[6] F. Xia, L. Sekaric, and Y. Vlasov. Ultra-compact optical buffers on a silicon chip. *Nature Photon.*, 1:65, 2007.

[7] Y. Vlasov, W. M. J. Green, and F. Xia. High-throughput silicon nanophotonic wavelength-insensitive switch for on-chip optical networks. *Nature Photon.*, 2:242, 2008.

[8] A. M. Armani, R. P. Kulkarni, S. E. Fraser, R. C. Flagan, and K. J. Vahala. Label-free, single-molecule detection with optical microcavities. *Science*, 317(5839):783, 2007.

[9] T. Carmon and K. J. Vahala. Visible continuous emission from a silica microphotonic device by third-harmonic generation. *Nature Phys.*, 3:430–435, 2007.

[10] T. Aoki, B. Dayan, E. Wilcut, W. P. Bowen, A. S. Parkins, T. J. Kippenberg, K. J. Vahala, and H. J. Kimble. Observation of strong coupling between one atom and a monolithic microresonator. *Nature*, 443:671–674, 2006.

[11] T. J. Kippenberg and K. J. Vahala. Cavity optomechanics: Back-action at the mesoscale. *Science*, 321:1172–1176, 2008.

[12] T. W. Ebbesen, H. J. Lezec, H. F. Ghaemi, T. Thio, and P. A. Wolff. Extraordinary optical transmission through sub-wavelength hole arrays. *Nature*, 391: 667–669, 1998.

[13] S. Nie and S. R. Emory. Probing single molecules and single nanoparticles by surface-enhanced raman scattering. *Science*, 275:1102–1106, 1997.

[14] S. I. Bozhevolnyi, V. S. Volkov, E. Devaux, J.-Y. Laluet, and T. W. Ebbesen. Channel plasmon subwavelength waveguide components including interferometers and ring resonators. *Nature*, 440:508–511, 2006.

[15] C. Enkrich, M. Wegener, S. Linden, S. Burger, L. Zschiedrich, F. Schmidt, J. F. Zhou, Th. Koschny, and C. M. Soukoulis. Magnetic metamaterials at telecommunication and visible frequencies. *Phys. Rev. Lett.*, 95(20):203901, 2005.

[16] G. Dolling, M. Wegener, C. M. Soukoulis, and S. Linden. Negative-index metamaterial at 780 nm wavelength. *Opt. Lett.*, 32(1):53–55, 2007.

[17] http://www.isiknowledge.com, search conducted on Dec 1, 2008.

[18] W. Reed and T. Hill. Triangular mesh methods for the neutron transport equation. Technical Report LA-UR-73-479, Los Alamos National Laboratory, 1973.

[19] J. S. Hesthaven and T. Warburton. Nodal high-order methods on unstructured grids - I. Time-domain solution of Maxwell's equations. *J. Comput. Phys.*, 181(1): 186–221, 2002.

[20] D. Jackson. *Classical Electrodynamics*. Wiley, 1999.

[21] R. Nevels and J. Jeong. The time domain Green's function and propagator for Maxwell's equations. *IEEE T. Antenn. Propag.*, 52(11):3012–3018, 2004.

[22] M. Galassi, J. Davies, J. Theiler, B. Gough, G. Jungman, M. Booth, and F. Rossi. *GNU Scientific Library: Reference Manual*. Network Theory Ltd., Bristol, UK, second revised edition, 2005.

[23] M. Born and E. Wolf. *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*. Cambridge University Press, Cambridge, England, seventh edition, October 1999.

[24] Tie-29: Refractive index and dispersion. Technical report, Schott AG, Germany, 2005.

[25] P. B. Johnson and R. W. Christy. Optical constants of the noble metals. *Phys. Rev. B*, 6(12):4370–4379, 1972.

[26] A. Vial and T. Laroche. Description of dispersion properties of metals by means of the critical points model and application to the study of resonant structures using the fdtd method. *J. Phys. D: Appl. Phys.*, 40:7152–7158, 2007.

[27] J.-P. Bérenger. *Perfectly Matched Layer (PML) for Computational Electromagnetics*. Morgan & Claypool, 2007.

[28] A. Taflove and S. C. Hagness. *Computational Electrodynamics: The Finite-Difference Time-Domain Method.* Artech House Publishers, third edition, 2005.

[29] F. L. Teixeira and W. C. Chew. General closed-form pml constitutive tensors to match arbitrary bianisotropic and dispersive linear media. *IEEE Microw. Guided W.*, 8:223–225, 1998.

[30] K. Yee. Numerical solution of initial boundary value problems involving maxwell's equations in isotropic media. *IEEE T. Antenn. Propag.*, 14:302–307, 1966.

[31] J. C. Strikwerda. *Finite Difference Schemes and Partial Differential Equations.* SIAM, 2004.

[32] B. Finkelstein and R. Kastner. Finite difference time domain dispersion reduction schemes. *J. Comput. Phys.*, 221(1):422–438, 2007.

[33] T. T. Zygiridis and T. D. Tsiboukis. Optimized three-dimensional fdtd discretizations of maxwell's equations on cartesian grids. *J. Comput. Phys.*, 226(2):2372–2388, 2007.

[34] A. Ditkowski, K. Dridi, and J. S. Hesthaven. Convergent cartesian grid methods for maxwell's equations in complex geometries. *J. Comput. Phys.*, 170(1):39–80, 2001.

[35] A. Farjadpour, D. Roundy, A. Rodriguez, M. Ibanescu, P. Bermel, J. D. Joannopoulos, S. G. Johnson, and G. Burr. Improving accuracy by subpixel smoothing in FDTD. *Opt. Lett.*, 31:2972–2974, 2006.

[36] J. L. Young and R. O. Nelson. A summary and systematic analysis of fdtd algorithms for linearly dispersive media. *IEEE Antennas Propag.*, 43(1), 2001.

[37] C. A. Balanis. *Advanced Engineering Electromagnetics.* Wiley, 1989.

[38] S. K. Lele. Compact finite difference schemes with spectral-like resolution. *J. Comput. Phys.*, 103:16–42, 1992.

[39] H. M. Jurgens and D. W. Zingg. Numerical solution of the time-domain maxwell equations using high-accuracy finite-difference methods. *SIAM J. Sci. Comput.*, 22(5):1675–1696, 2001.

[40] J. S. Hesthaven and T. Warburton. *Nodal Discontinuous Galerkin Methods.* Springer, 2007.

[41] R. L. LeVeque. *Finite Volume Methods for Hyperbolic Problems.* Cambridge University Press, 2002.

[42] E. F. Toro. *Riemann Solvers and Numerical Methods for Fluid Dynamics.* Springer, 1999.

[43] L. Komzsik. *Approximation Techniques for Engineers*. CRC Press, 2006.

[44] J. Schöberl. NETGEN an advancing front 2d/3d-mesh generator based on abstract rules. *Comput. Visual Sci.*, 1(1):41–52, 1997.

[45] P. Monk. *Finite Element Methods for Maxwell's Equations*. Oxford University Press, 2003.

[46] J. S. Hesthaven. From electrostatics to almost optimal nodal sets for polynomial interpolation in a simplex. *SIAM J. Numer. Anal.*, 35(2):655–676, 1998.

[47] T. Warburton. An explicit construction of interpolation nodes on the simplex. *J. Eng. Math.*, 56(3):247–262, 2006.

[48] W. Heinrichs. Improved lebesgue constants on the triangle. *J. Comput. Phys.*, 207(2):625–638, 2005.

[49] J. H. Williamson. Low-storage runge-kutta schemes. *J. Comput. Phys.*, 35(1): 48–56, 1980.

[50] M. H. Carpenter and C. A. Kennedy. Fourth-Order 2N-Storage Runge-Kutta Schemes. Technical Report NASA-TM-109112, NASA Langley Research Center, VA, JUN 1994.

[51] T. Lu, P. W. Zhang, and W. Cai. Discontinuous Galerkin methods for dispersive and lossy Maxwell's equations and PML boundary conditions. *J. Comput. Phys.*, 200(2):549–580, 2004.

[52] J. Niegemann, M. König, K. Stannigel, and K. Busch. Higher-order time-domain methods for the analysis of nano-photonic systems. *Photon. Nanostruct.: Fundam. Appl.*, 2008. (in press).

[53] C. Moler and C. van Loan. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Review*, 45(1):3–49, 2003.

[54] H. A. De Raedt, J. S. Kole, K. F. L. Michielsen, and M. T. Figge. Unified framework for numerical methods for solving the time-dependent maxwell equations. *Comp. Phys. Comm.*, 156:43–61, 2003.

[55] M. Pototschnig, J. Niegemann, L. Tkeshelashvili, and K. Busch. Time-domain simulations of the nonlinear maxwell equations using operator-exponential methods. *IEEE T. Antenn. Propag.*, 2009. (in press).

[56] Y. Saad. Analysis of some krylov subspace approximations to the matrix exponential operator. *SIAM J. Numer. Anal.*, 29(1):209–228, 1992.

[57] W. E. Arnoldi. The principle of minimized iteration in the solution of the matrix eigenvalue problem. *Quart. Appl. Math.*, 9:17, 1951.

[58] E. Gallopoulos and Y. Saad. Efficient solution of parabolic equations by krylov approximation methods. *SIAM J. Sci. Stat. Comput.*, 13(5):1236–1264, 1992.

[59] R. B. Sidje. EXPOKIT: Software package for computing matrix exponentials. *ACM Trans. Math. Softw.*, 24:130, 1998.

[60] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery. *Numerical Recipes in C*. Cambridge University Press, Cambridge, UK, 2nd edition, 1992.

[61] A. A. Asatryan, K. Busch, R. C. McPhedran, L. C. Botton, C. M. de Sterke, and N. A. Nicorovici. Two-dimensional green tensor and local density of states in finite-sized two-dimensional photonic crystals. *Wave in Random Media*, 13:9, 2003.

[62] C. Fabry and A. Perot. Théorie et applications d'une nouvelle méthode de spectroscopie interférentielle. *Ann. Chim. Phys.*, 16:115, 1899.

[63] K. J. Vahala. Optical microresonators. *Nature*, 424(8):839–846, 2003.

[64] V. A. Mandelshtam and H. S. Taylor. Harmonic inversion of time signals and its applications. *J. Chem. Phys.*, 107(17):6756–6769, 1997.

[65] S. G. Johnson. `http://ab-initio.mit.edu/wiki/index.php/Harminv`.

[66] G. Mie. Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen. *Ann. Phys.*, 330(3):377–445, 1908.

[67] C. F. Bohren and D. R. Huffman. *Absorption and scattering of light by small particles*. John Wiley & Sons, New York, USA, 2nd edition, 1983.

[68] H. C. Van de Hulst. *Light Scattering by Small Particles*. Dover, 2nd edition, 1981.

[69] J. A. Stratton. *Electromagnetic Theory*. McGraw-Hill, New York, USA, 1941.

[70] S. C. Hagness, D. Rafizadeh, S. T. Ho, and A. Taflove. FDTD microcavity simulations: Design and experimental realization of waveguide-coupled single-mode ring and whispering-gallery-mode disk resonators. *J. Lightwave Technol.*, 15:2154–2165, 1997.

[71] M. Borselli, K. Srinivasan, P. E. Barclay, and O. Painter. Rayleigh scattering, mode coupling, and optical loss in silicon microdisks. *Appl. Phys. Lett.*, 85(17): 3693–3695, 2004.

[72] G. Lenz, B. J. Eggleton, C. K. Madsen, and R. E. Slusher. Optical delay lines based on optical filters. *IEEE J. Quantum Elect.*, 37:525, 2001.

[73] O. Schwelb. Transmission, group delay, and dispersion in single-ring optical resonators and add/drop filters - a tutorial overview. *J. Lightwave Tech.*, 22:1380, 2004.

[74] W. Pernice, 2008. private communication.

[75] Y. Takakura. Optical resonance in a narrow slit in a thick metallic screen. *Phys. Rev. Lett.*, 86(24):5601–5603, 2001.

[76] F. J. García-Vidal, H. J. Lezec, T. W. Ebbesen, and L. Martín-Moreno. Multiple paths to enhance optical transmission through a single subwavelength slit. *Phys. Rev. Lett.*, 90(21):213901, 2003.

[77] H. Liu and P. Lalanne. Microscopic theory of the extraordinary optical transmission. *Nature*, 452:728–731, 2008.

[78] G. Dolling, Ch. Enkrich, M. Wegener, C. M. Soukoulis, and S. Linden. Low-loss negative-index metamaterial at telecommunication wavelengths. *Opt. Lett.*, 31 (12):1800–1802, 2006.

[79] B. J. Wiley, Y. Chen, J. M. McLellan, Y. Xiong, Z.-Y. Li, D. Ginger, and Y. Xia. Synthesis and optical properties of silver nanobars and nanorice. *Nano Lett.*, 7 (4):1032–1036, 2007.

[80] F. Wang and Y. Ron Shen. General properties of local plasmons in metal nanostructures. *Phys. Rev. Lett.*, 97(20):206806, 2006.

[81] M. I. Stockman, S. V. Faleev, and D. J. Bergman. Coherent control of femtosecond energy localization in nanosystems. *Phys. Rev. Lett.*, 88(6):067402, 2002.

[82] M. I. Stockman, D. J. Bergman, and T. Kobayashi. Coherent control of nanoscale localization of ultrafast optical excitation in nanosystems. *Phys. Rev. B*, 69(5): 054202, 2004.

[83] K. Stannigel. Modeling the linear and nonlinear properties of metallic nanostructures using discontinuous galerkin finite element time-domain methods. Master's thesis, Universität Karlsruhe, 2008.

Acknowledgements

Finally, I would like to acknowledge a number of people who all contributed to this thesis in one way or another.

My utmost gratitude goes to my thesis advisor, Prof. Dr. Kurt Busch for his expertise, constant support and trust. He gave me the freedom and time to explore my own ideas, yet he was available when needed. I would also like to thank Prof. Dr. Martin Wegener for agreeing to co-referee this thesis.

I am deeply indebted to my former masters students Martin Pototschnig, Michael König, Jan Gieseler and Kai Stannigel for their scientific contributions. A lot of their work has directly or indirectly entered this thesis and many of the results I presented would not have been available without them.

In addition, I am very grateful to Daniel Hermann, Sabine Essig, Patrick Mack, Michael König, Kai Stannigel and Jan Gieseler for proof-reading parts of this thesis.

Especially, I would like to thank the members of the first generation of this group, Frank Hagmann, Daniel Hermann, Meikel Frank, Matthias Schillinger, Marcus Diem and Lasha Tkeshelashvili, for kindly welcoming me into the group. Furthermore, I thank all the current and former members of the group for the great working atmosphere. The countless stimulating discussions, the regular barbecues, and the endless supply of cake made the past years an incredibly pleasant experience.

I can not end without thanking Franziska for her constant support and encouragement. It is to her that I dedicate this thesis.

List of Publications

1. L. Tkeshelashvili, J. Niegemann und K. Busch "*Nonlinear Wave Interaction In Photonic Band Gap Material*", Photonics and Nanostructures—Fundamentals and Applications **4**, 75 (2006)

2. J. Niegemann, L. Tkeshelashvili und K. Busch "*Higher-Order Time-Domain Simulations of Maxwell's Equations using Krylov-Subspace Methods*", Journal of Computational and Theoretical Nanoscience **4**, 627 (2007)

3. K. Busch, J. Niegemann, M. Pototschnig und L. Tkeshelashvili "*A Krylov-Subspace Based Solver For The Linear And Nonlinear Maxwell Equations*", Physica Status Solidi **B 244**, 3479 (2007)

4. S. Essig, J. Niegemann, L. Tkeshelashvili und K. Busch "*Solitary Wave Formation in One- Dimensional Photonic Crystals*" , Physica Status Solidi **A 204**, 3591 (2007)

5. J. Niegemann, L. Tkeshelashvili und K. Busch "*Chaotic Scattering Of Solitons On Point Defects In Fiber Bragg Gratings*", Optics Express **16**, 110170 (2008)

6. M. Husnik, M. W. Klein, N. Feth, M. König, J. Niegemann, K. Busch, S. Linden und M. Wegener "*Absolute Extinction Cross-Section Of Individual Split-Ring Resonators*", Nature Photonics **2**, 614 (2008)

7. J. Niegemann, M. König, K. Stannigel und K. Busch "*Higher-Order Time-Domain Methods for the Analysis of Nano-Photonic Systems*", Photonics and Nanostructures—Fundamentals and Applications **7**, 2 (2009)

8. M. Pototschnig, J. Niegemann, L. Tkeshelashvili und K. Busch "*Time-Domain Simulations of the Nonlinear Maxwell Equations Using Operator-Exponential Methods*", IEEE Transactions on Antennas and Propagation **57**, 475 (2009)