The Full Anisotropic Adaptive Fourier Modal Method and its Application to Periodic and Aperiodic Photonic Nanostructures

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

von der Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

Dipl.-Phys. Thomas Zebrowski
aus Baden-Baden

Tag der mündlichen Prüfung: 19. Oktober 2012
Referent: Prof. Dr. Kurt Busch
Korreferent: Prof. Dr. Martin Wegener

Contents

Chapter 1.

1 Introduction

Visible light captured by the eye is the most important sense to human beings. It enables us to remotely perceive the world around us, close by as well as parsecs away.

Light transports enormous amounts of information. In daily life it tells us about the dimensions and distance of objects, their structure or surface texture, and their constituent materials. With its large propagation velocity — a fundamental natural constant — it enables us to gather information momentarily and also look into the universe's past. With optical tools like telescopes we can learn about large scale objects such as galaxies, with microscopes we can analyze tiny objects like cells. Light mediates the knowledge stored on computers somewhere at the other end of the earth, and the knowledge contained in books on our desk. Light transports energy.

Light is prerequisite. It is appealing, fascinating, and important to humans. We rely on it. In absence it is missed. This is the reason why tremendous scientific efforts were spend to gain control over it, to be able to create and manipulate it. Scientific progress as well as improved technical skills and machinery facilitated a deeper understanding of its nature troughout history, but particularly within the last decades. Today, we know that light is created by matter. The information it transports is picked up by interaction with matter. Thus, the control of light must happen by control of the matter on the visible light's inherent scale — the nano-scale.

Among the groundbreaking recent advancements is clearly the proposal of photonic crystals (PCs) by Yablonovitch [1] and John [2] in 1986. They were first to realize that a regular, periodic pattern of different dielectric materials leads to a photonic band structure which can feature complete photonic bandgaps (PBG) similar to electronic band structures in semiconductors. These complete bandgaps are frequency regions where light propagation within the crystal is forbidden irrespective of the propagation direction. This finding stimulated enormous endeavors to physically realize such artificial materials and selectively engineer their properties. While at first production techniques allowed only for high precision structuring in the micrometer regime [3], the structured materials' feature sizes in subsequent works eventually advanced into the high nanometer regime such that the bandgap appeared in the infrared part of the spectrum [4, 5].

One setup prooved particularly dependable: The woodpile photonic crystal structure. With the help of direct laser writing (DLW), a newly developed structuring technique for rapid prototyping, it became possible to reach in-plane rod spacings well below one micrometer and, with silicon as rod material, a bandgap in the near infrared [6]. However, a complete photonic bandgap in the visible

was still out of reach and remained challenging. First of all, this would require even smaller feature sizes, and secondly a high refractive index contrast between rod and surrounding material constituted from materials with low absorption in the visible range.

Lately, stimulated emission depletion direct laser writing (STED-DLW) [7,8] and a newly developed titania (TiO_2) atomic layer deposition (ALD) double inversion process — for the conversion of the polymeric template into the final material — enabled the realization of woodpiles with the required characteristics for a complete photonic bandgap in the visible [9]. The geometric parameters deduced from scanning electron microscope (SEM) pictures indicate that the realized geometry is compatible with designs which exhibit a complete bandgap. While these indications were obtained with an established numerical bandstructure simulation tool [10], there remains the usual gap in the chain of evidence. Any bandstructure simulation naturally considers an infinitely (three-dimensional) periodic bulk system. Every realized photonic crystal, however, is of finite dimensions and features a structure surface. Hence, the experimentally available meassurements obtained from angle- and polarization-resolved transmittance spectroscopy (cf. [11]), must be savely linked to the bandstructure results. As part of this work, we will provide this link and strong evidence for the successful realization of the complete three-dimensional photonic bandgap in the visible — employing the perfectly suited Fourier modal method (FMM).

Early on in the scientific exploration of PCs, the focus was not only on the realization of perfectly regular bulk structures, but also on the principle's application to useful devices. One of the directions of development was the utilization of the electromagnetic isolation effect PCs provide in the frequency range of the bandgap. Introducing an imperfection into the otherwise perfect pattern of such structures leads to a local confinement of the light in the defect's vicinity. This effect was, for instance, adopted to optical fibers and opened up the field of photonic crystal fibers (PCF) [12, 13].

If the photonic crystal part consists of materials whose optical properties can be specifically manipulated, the bandgap can be tuned [14–17]. One sort of such materials are for example anisotropic liquid crystals (LC) controled by an external electric field. Infiltrated into a PCF and enclosed by a periodic electrode along the fiber axis, they can be used to dynamically induce a long period grating (LPG) [18]. These LPGs are interesting devices for dynamical fiber based filter or switching applications. Due to their huge length in the order of several millimeters in comparison to their small diameter of a few microns, these systems are challenging for numerical simulation tools.

An eigenmode expansion based numerical method like the FMM is one of the most promising candidates for an accurate simulation of such devices, because it features a good scaling behavior with respect to the system size along the fiber axis. However, the FMM is optimized to handle infinitely periodic systems in the plane transverse to the fiber axis, like photonic crystals. Hence, the major part of this work is dedicated to the optimization of the method towards fiber-based photonic systems such as the LPG. This comprises fully anisotropic material tensors, adapted coordinate (AC) meshes, an adapted spatial resolution (ASR), and perfectly matched layer (PML) boundary conditions.

Overview

We commence in Chap. 2 with the discussion of Maxwell's equations as the underlying physical laws for the description of light and its propagation. Starting from their common form, we derive step by step a form suitable for a numerical treatment of problems in the frequency domain. Subsequently, we give a short digest of field continuity conditions at material interfaces and energy transport. Plane

waves are introduced as solutions to Maxwell's equations in homogenous space. Furthermore, we give a detailed introduction into the derivation of a formulation suitable for arbitrary curvilinear coordinate systems — the covariant formulation.

The subsequent chapter, Chap. 3, familiarizes the reader with the mathematical description of periodic systems by means of the concepts of lattice and reciprocal lattice. We shortly recapitulate Bloch modes as their natural solutions, before we devote a large section to numerical discretization and description of the problem in Fourier space. In the last part, we acquaint ourselves with the theory of diffraction.

Chapter 4 gives an introduction into the field of waveguides and the properties of waveguide modes. Since the circular step-index fiber is one of the few systems with available analytical solutions, we adopt them for later use as reference solutions.

The abstract scheme of modal methods is developed in Chap. 5. Besides the systems decomposition into layers and the eigenmode expansion, the major part of the chapter is devoted to the scattering matrix algorithm as integral part of all modal methods. Serveral different field matching schemes are delt with as well as all kinds of procedures and usefull tricks.

The Fourier modal method as our particular variant of modal methods for periodic systems is detailed in Chap. 6. We present the full anisotropic eigenvalue problem in the discretized form and cover all aspects like field sources, transmittance and reflectance as calculated quantities, and the reconstruction of the field solutions with different approaches. Last, we give a short introduction into symmetry reductions at the example of the C_{2v} symmetry.

Chapter 7 is concerned with real and complex coordinate transformations as extensions to the ordinary FMM — not solely, but also — towards the treatment of fiber systems. This application of transformation optics concepts strongly grounds on the covariant formulation of Maxwell's equations introduced in Chap. 2. Intuitive construction schemes for the creation of real coordinate transformations are developed with the aim of an improved surface representation by adaptive coordinates (AC) and an adaptive spatial resolution (ASR) of discontinuities in the material functions. Complex coordinate transformations are exploited for the electromagnetic isolation of aperiodic structures by perfectly matched layers (PML), so that the FMM can handle them as well. We discuss advantages and disadvantages of the presented techniques and different strategies for their integration into the method.

The newly implemented extensions are rigorously tested and validated in Chap. 8. We evaluate the performance of different mesh types as well as different PML types. Our attention is particularly focussed on the discussion of obstacles and issues and how they can be accounted for.

Last not least, we apply the developed numerical simulation code to challenging and exciting applications in Chap. 9. The first system, the woodpile photonic crystal with a complete bandgap in the visible, is a periodic system and, thus, treated with classical FMM. All developed extensions of the FMM — full anisotropy, adaptive coordinates, adaptive spatial resolution, and perfectly matched layers — find their successful application in the design and simulation of a LC infiltrated fiber based LPG mode coupler.

The thesis is concluded with a summary and an outlook.

Chapter 2.
2 Fundamentals of Electrodynamics

Light-matter interaction in nanophotonic systems is described by the fundamental equations of electrodynamics. In the physical systems we would like to investigate, individual quantum effects do not play an observable role. Rather the averaged material properties in interplay with light-fields dominate the physical processes. These processes are best treated in the framework of classical electrodynamics.

In this chapter, we introduce the reader into the topic of classical electromagnetics and establish the theoretical background for the remainder of this work. We start with the introduction of electromagnetic fields and the presentation of Maxwell's equations in Sec. 2.1. Maxwell's equations formulate the physical laws that govern the examined interaction between fields and matter. We gradually reformulate and simplify the equations to ease a numerical treatment. Section 2.2 is concerned with the continuity of fields at material interfaces, and Sec. 2.3 covers the transport of energy contained in the fields. The wave equation and propagation phenomena of plane waves are briefly discussed in Sec. 2.4. In the end of this chapter, in Sec. 2.5, we give a profound introduction into curvilinear coordinate systems and show that Maxwell's equations can be formulated in such a way that they are invariant independent of the palpable chosen coordinate system.

Throughout this chapter, we follow the derivations of the electrodynamics textbook written by J. D. Jackson [19] and the solid state textbooks written by Ch. Kittel [20] and N. W. Ashcroft [21]. The disquisition on curvilinear coordinates is oriented along the path laid out in the work of Leonhardt and Philbin [22].

2.1. Maxwell's Equations

The numerical method we develop and use in the course of this thesis builds up on the fundamental equations established by James C. Maxwell in 1865 in his paper "A Dynamical Theory of the Electromagnetic Field" [23]. Using the International System of Units (SI) and the notation coined by O.

Heavyside, Maxwell's equations read

$$\boldsymbol{\nabla} \cdot \mathbf{D}(\mathbf{r}, t) = \rho_{\mathrm{f}}(\mathbf{r}, t)\,, \tag{2.1a}$$
$$\boldsymbol{\nabla} \cdot \mathbf{B}(\mathbf{r}, t) = 0\,, \tag{2.1b}$$
$$\boldsymbol{\nabla} \times \mathbf{E}(\mathbf{r}, t) = -\partial_t \mathbf{B}(\mathbf{r}, t)\,, \tag{2.1c}$$
$$\boldsymbol{\nabla} \times \mathbf{H}(\mathbf{r}, t) = \mathbf{j}_{\mathrm{f}}(\mathbf{r}, t) + \partial_t \mathbf{D}(\mathbf{r}, t)\,. \tag{2.1d}$$

These equations describe the interplay between electric field $\mathbf{E}$, magnetic field $\mathbf{H}$, electric displacement field $\mathbf{D}$, and magnetic induction field $\mathbf{B}$. Electric charges and currents which are not bound to a medium are represented by the free electric current density $\mathbf{j}_{\mathrm{f}}$ and free electric charges ρ_{f}. They serve as sources for the electromagnetic fields. All introduced quantities depend on the spatial coordinate $\mathbf{r} = (x, y, z)^T$ and time t.

Maxwell's equations reflect the physical laws that fully govern electromagnetic effects on a macroscopic length scale. Macroscopic thereby refers to a regime a good deal larger then the atomic length scales, where quantum mechanical effects play a major role. In this regime the atomic physical processes are subsumed in spatially averaged effective media properties and the fields are to be considered correspondingly. These effective media properties enter via the electric displacement field $\mathbf{D}$ and magnetic induction field $\mathbf{B}$ and the constitutive relations introduced in Sec. 2.1.3.

Of all four fields only the electric field $\mathbf{E}$ and the magnetic induction field $\mathbf{B}$ are directly observable via the Lorentz force

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})\,, \tag{2.2}$$

that acts on an electrical charge q moving with velocity $\mathbf{v}$.

The focus of this thesis lies on physical aspects much less general than what Maxwell's above stated equations account for. The following paragraphs therefore introduce simplifications and specializations leading to the form of Maxwell's equations which will be the starting point of the numerical methods introduced in Chap. 5 and Chap. 6.

2.1.1. Sources

In this thesis we are not so much interested in the creation of electromagnetic fields. We rather study the evolution of a given initial field in space and time for a given structure made of solid, liquid or gaseous matter. Hence, only systems with $\rho_{\mathrm{f}}(\mathbf{r}, t) = 0$ and $\mathbf{j}_{\mathrm{f}}(\mathbf{r}, t) = \mathbf{0}$ will be considered and the respective terms in Eq. (2.1a) and Eq. (2.1d) can be neglected.

2.1.2. Frequency Domain

Our focus will additionally be on time harmonic fields and stationary solutions of Maxwell's equations. So, we make the ansatz

$$\boldsymbol{\Psi}(\mathbf{r}, t) = \boldsymbol{\Psi}(\mathbf{r}, \omega)\, e^{-\mathrm{i}\omega t} \tag{2.3}$$

where $\boldsymbol{\Psi}$ is a place holder for all electromagnetic fields and ω is the (angular) frequency[1]. Consequently, the time derivatives in Eq. (2.1c) and Eq. (2.1d) can be replaced by $\partial_t \rightarrow -\mathrm{i}\omega$. The

[1] Throughout this thesis we use frequency synonymous to angular frequency.

exponential time dependence can and will be omitted in the presentation of the equations as well as in the remainder of the thesis, but is always implicitly assumed. All fields then depend on space $\mathbf{r}$ and frequency ω; this representation we call frequency domain. Maxwell's source free equations in frequency domain read

$$\boldsymbol{\nabla} \cdot \mathbf{D}(\mathbf{r},\omega) = 0\,, \tag{2.4a}$$
$$\boldsymbol{\nabla} \cdot \mathbf{B}(\mathbf{r},\omega) = 0\,, \tag{2.4b}$$
$$\boldsymbol{\nabla} \times \mathbf{E}(\mathbf{r},\omega) = \quad \mathrm{i}\omega\, \mathbf{B}(\mathbf{r},\omega)\,, \tag{2.4c}$$
$$\boldsymbol{\nabla} \times \mathbf{H}(\mathbf{r},\omega) = -\mathrm{i}\omega\, \mathbf{D}(\mathbf{r},\omega)\,. \tag{2.4d}$$

2.1.3. Constitutive Relations

So far, Maxwell's equations in the frequency domain are not unambiguously solvable. The eight equations contain twelve independent field components. The missing equations are provided by the constitutive relations

$$\mathbf{D} = \mathbf{D}(\mathbf{E},\mathbf{H})\,, \qquad \mathbf{B} = \mathbf{B}(\mathbf{E},\mathbf{H})\,. \tag{2.5}$$

They are called constitutive or material relations because they describe the material response due to the external electromagnetic stimulation. This dependence is in general complicated. However, in the most common optical cases it is sufficient to describe the response by the polarization $\mathbf{P}$ and the magnetization $\mathbf{M}$ of the material and neglect higher order moments. Furthermore, in the considered cases $\mathbf{P}$ usually depends much stronger on $\mathbf{E}$ than on $\mathbf{H}$. The same also applies to $\mathbf{M}$, but the other way round. These limitations lead to

$$\mathbf{D}(\mathbf{r},\omega) = \varepsilon_0 \mathbf{E}(\mathbf{r},\omega) + \mathbf{P}(\mathbf{E}(\mathbf{r},\omega))\,, \tag{2.6a}$$
$$\mathbf{B}(\mathbf{r},\omega) = \mu_0 \mathbf{H}(\mathbf{r},\omega) + \mathbf{M}(\mathbf{H}(\mathbf{r},\omega))\,, \tag{2.6b}$$

where the first terms on the right hand side constitute the respective vacuum contributions[2].

For local media polarization and magnetization can be expanded into a power series in the electric and magnetic field, respectively. Since we are interested in electromagnetic field strengths that are small compared to atomic fields, only the linear term provides a significant contribution and we restrict our considerations to such linear materials. Hence, polarization and magnetization can be expressed by

$$\mathbf{P}(\mathbf{r},\omega) = \varepsilon_0\, \underline{\boldsymbol{\chi}}_e^{(1)}(\mathbf{r},\omega)\; \mathbf{E}(\mathbf{r},\omega)\,, \tag{2.7a}$$
$$\mathbf{M}(\mathbf{r},\omega) = \mu_0\, \underline{\boldsymbol{\chi}}_m^{(1)}(\mathbf{r},\omega)\, \mathbf{H}(\mathbf{r},\omega)\,, \tag{2.7b}$$

with the tensorial first order expansion coefficients called electric susceptibility $\underline{\boldsymbol{\chi}}_e$ and magnetic susceptibility $\underline{\boldsymbol{\chi}}_m$. Inserting Eqs. (2.7) into Eqs. (2.6) and introducing the permittivity[3] $\underline{\boldsymbol{\varepsilon}} = (1+\underline{\boldsymbol{\chi}}_e)$ and the permeability $\underline{\boldsymbol{\mu}} = (1+\underline{\boldsymbol{\chi}}_m)$, the linear constitutive relations read

$$\mathbf{D}(\mathbf{r},\omega) = \varepsilon_0\, \underline{\boldsymbol{\varepsilon}}(\mathbf{r},\omega)\; \mathbf{E}(\mathbf{r},\omega)\,, \tag{2.8a}$$
$$\mathbf{B}(\mathbf{r},\omega) = \mu_0\, \underline{\boldsymbol{\mu}}(\mathbf{r},\omega)\, \mathbf{H}(\mathbf{r},\omega)\,. \tag{2.8b}$$

[2] The used natural constants are defined as: $\varepsilon_0 = 1/(\mu_0\, c_0^2)$ is the vacuum permittivity, $\mu_0 = 4\pi \cdot 10^{-7}\ \mathrm{Vs/(Am)}$ is the vacuum permittivity, and $c_0 = 299\,792\,458\ \mathrm{m/s}$ the vacuum speed of light.

[3] To be precise, this is the relative permittivity and relative permeability. However, the "relative" will be omitted throughout the thesis.

The permittivity and permeability are both second rank tensors (3×3) which allow for the treatment of anisotropic materials. Their frequency dependency describes dispersion. For many dielectric materials both characteristics are not important and the permittivity can be reduced to the scalar dielectric constant ϵ. The permeability μ at optical frequencies usually equals one. Despite these facts, we would like to emphasize that for the course of this thesis these characteristics are important and will be used. Therefore, we carry them through all further derivations.

The restriction to linear constitutive relations entails that Maxwell's equations become linear partial differential equations. As such, any superposition of their solutions remains a solution itself. In particular, with the Fourier transform

$$\mathbf{\Psi}(\mathbf{r}, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}\omega \; \mathbf{\Psi}(\mathbf{r}, \omega) \; e^{-\mathrm{i}\omega t} \tag{2.9}$$

and its inverse it is possible to calculate arbitrary time dependent signals as well and it becomes clear that our ansatz from Eq. (2.3) is simply a special case. Nevertheless, this special case is the most commonly used.

2.1.4. Reduction to Curl Equations

The constitutive relations derived in Sec. 2.1.3 reduce the unknowns in Maxwell's equations to six. With all eight of Maxwell's equations the system is overdetermined. Thus, as long as our solutions fulfill the two divergence equations, Eq. (2.1a) and Eq. (2.1b), it is sufficient to keep the six curl equations Eq. (2.1c) and Eq. (2.1d) in order to solve the whole system of equations.

When the simplified curl equations in frequency domain are solved for

$$\mathbf{B} = \frac{1}{\mathrm{i}\omega} \boldsymbol{\nabla} \times \mathbf{E} \qquad \text{and} \qquad \mathbf{D} = \frac{-1}{\mathrm{i}\omega} \boldsymbol{\nabla} \times \mathbf{H} \,, \tag{2.10}$$

it becomes apparent that because of the identity relation $\boldsymbol{\nabla} \cdot (\boldsymbol{\nabla} \times \mathbf{\Psi}) = 0$ their solutions are automatically divergence free ($\rho_\mathrm{f} = 0$). Hence, we content ourselves to work with the curl equations solely.

2.1.5. Dimensionless Units and Fields

Maxwell's equations, Eqs. (2.1), with all reductions and simplifications applied in Sec. 2.1.1 to Sec. 2.1.4, read in the SI system

$$\boldsymbol{\nabla}_\mathrm{SI} \times \mathbf{E}_\mathrm{SI}(\mathbf{r}_\mathrm{SI}, \omega_\mathrm{SI}) = \quad \mathrm{i}\omega_\mathrm{SI} \, \mu_0 \, \underline{\boldsymbol{\mu}}(\mathbf{r}_\mathrm{SI}, \omega_\mathrm{SI}) \, \mathbf{H}_\mathrm{SI}(\mathbf{r}_\mathrm{SI}, \omega_\mathrm{SI}) \,, \tag{2.11a}$$

$$\boldsymbol{\nabla}_\mathrm{SI} \times \mathbf{H}_\mathrm{SI}(\mathbf{r}_\mathrm{SI}, \omega_\mathrm{SI}) = -\mathrm{i}\omega_\mathrm{SI} \, \varepsilon_0 \, \underline{\boldsymbol{\varepsilon}}(\mathbf{r}_\mathrm{SI}, \omega_\mathrm{SI}) \, \mathbf{E}_\mathrm{SI}(\mathbf{r}_\mathrm{SI}, \omega_\mathrm{SI}) \,. \tag{2.11b}$$

Most of the quantities expressed here are still afflicted with a variety of different units. To prepare the equations for numerical evaluation we have to eliminate them and express the equations with dimensionless numbers. To this end, the SI quantities are split into a number and a scaling factor containing the units. The position vector $\mathbf{r}_\mathrm{SI} = \mathbf{r} \cdot a$ is split into the dimensionless vector $\mathbf{r}$ and the dimension tainted length scaling factor a, for instance. The length scaling factor is usually

a characteristic length of the investigated system, for example, the lattice constant of a periodic structure and typically in the order of $a = 1\,\mu\mathrm{m}$. With the help of a and the vacuum speed of light c_0 all remaining quantities except the fields can be treated the same way. Table 2.1 summarizes the relations for the conversion between SI and dimensionless units.

Quantity in Dimensionless Units	**In SI Units**
Position $\mathbf{r}$	$\mathbf{r}_{\mathrm{SI}} = \mathbf{r} \cdot a$
Wavelength λ	$\lambda_{\mathrm{SI}} = \lambda \cdot a$
Wave vector $\mathbf{k}$	$\mathbf{k}_{\mathrm{SI}} = \mathbf{k} \cdot \frac{1}{a}$
Time t	$t_{\mathrm{SI}} = t \cdot \frac{a}{c_0}$
Frequency ω	$\omega_{\mathrm{SI}} = \omega \cdot \frac{c_0}{a}$
Velocity $\mathbf{v}$	$\mathbf{v}_{\mathrm{SI}} = \mathbf{v} \cdot c_0$
Spatial Derivative $\boldsymbol{\nabla}$	$\boldsymbol{\nabla}_{\mathrm{SI}} = \boldsymbol{\nabla} \cdot \frac{1}{a}$
Electric Field $\mathbf{E}(\mathbf{r}, \omega)$	$\mathbf{E}_{\mathrm{SI}}(\mathbf{r}, \omega) = \mathbf{E}(\mathbf{r}, \omega) \cdot \frac{1}{\omega c_0 \varepsilon_0} H_0$
Magnetic Field $\mathbf{H}(\mathbf{r}, \omega)$	$\mathbf{H}_{\mathrm{SI}}(\mathbf{r}, \omega) = \mathbf{H}(\mathbf{r}, \omega) \cdot H_0$

Table 2.1.: *Dimensionless quantities and their relations to SI quantities.*

The electromagnetic fields need an additional scaling factor, e.g. $\mathbf{H}_{\mathrm{SI}}(\mathbf{r}, \omega) = \mathbf{H}(\mathbf{r}, \omega) \cdot H_0$. Here, again, the vector $\mathbf{H}$ is dimensionless and the field scaling factor is in optics typically in the order of $H_0 \approx 10^{-3}\,\mathrm{A/m}$. The electric field then scales[4] like mentioned in Tab. 2.1, where we employed the identities

$$c_0 = \frac{1}{\sqrt{\varepsilon_0 \mu_0}}\,, \qquad c_0 \mu_0 = \frac{1}{c_0 \varepsilon_0} = \frac{\sqrt{\mu_0}}{\sqrt{\varepsilon_0}}\,. \tag{2.12}$$

Replacing SI quantities in Eqs. (2.11) by the dimensionless quantities from Tab. 2.1 and dividing out the field scaling factor H_0 finally gives Maxwell's equations in dimensionless units

$$\boldsymbol{\nabla} \times \mathbf{E}(\mathbf{r}, \omega) = \quad \mathrm{i}\omega^2 \underline{\boldsymbol{\mu}}(\mathbf{r}, \omega)\, \mathbf{H}(\mathbf{r}, \omega)\,, \tag{2.13a}$$

$$\boldsymbol{\nabla} \times \mathbf{H}(\mathbf{r}, \omega) = -\mathrm{i} \quad \underline{\boldsymbol{\varepsilon}}(\mathbf{r}, \omega)\, \mathbf{E}(\mathbf{r}, \omega)\,. \tag{2.13b}$$

Note that also the arguments need to be replaced, but this is usually done in the following way: Before the evaluation of the equations the analyzed structure described by the permittivity $\underline{\boldsymbol{\varepsilon}}(\mathbf{r}, \omega)$ is directly defined in dimensionless units. After the evaluation, if necessary, the fields are transformed back to SI units as defined in Tab. 2.1. Moreover, by multiplication of the appropriate scaling factors to the arguments, we obtain

$$\mathbf{H}_{\mathrm{SI}}(\mathbf{r}_{\mathrm{SI}}, \omega_{\mathrm{SI}}) \overset{\text{Tab. 2.1}}{=} \mathbf{H}(a\mathbf{r},\, \omega c_0/a) \cdot H_0 \quad \overset{a,\, H_0}{\Longleftrightarrow} \quad \mathbf{H}(\mathbf{r}, \omega)\,. \tag{2.14}$$

The freedom in choice of the scaling factors without altering the form of Maxwell's equations Eqs. (2.13) is called scale invariance. It means that all physical solutions of the dimensionless equations that can be obtained by specifying a and H_0 are equivalent.

[4] A particularity of our choice for $\mathbf{E}_{\mathrm{SI}}$ is the additional dimensionless term $1/\omega$ which is not necessary. However, our implementations involve this factor for historic reasons [24].

As mentioned before, Maxwell's equations in the form of Eqs. (2.13) are the starting point for the description of the numerical algorithms used for their evaluation. These algorithms will be introduced in Chap. 5 and Chap. 6. But before we get there, there are a few more fundamental optical properties that have to be introduced first.

2.2. Continuity Conditions

Electromagnetic fields in homogeneous or continuously varying media are continuous themselves. Many interesting optical phenomena, however, are crucially connected to interfaces between different media where the material parameters as functions of position are discontinuous. The objective of this section is to derive from Maxwell's equations how the electromagnetic fields behave, whether they follow the discontinuity of the material parameters or not.

Thereto, we start out with Maxwell's source free divergence equations Eq. (2.4a) and Eq. (2.4b) in frequency domain and integrate them over a volume V enclosing part of the material interface – a so called Gaussian pillbox – as depicted in Fig. 2.1. The top and bottom faces of the pillbox are parallel to the material interface and of base area $\Delta a \neq 0$ each. Its sidewalls are of height δ. Application of

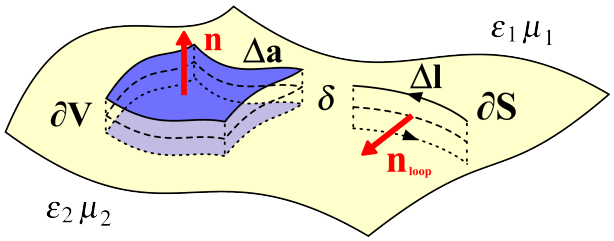


Figure 2.1.: *Integration contours at a material interface. Left: Gaussian pillbox. Right: Stokesian loop.*

the Gauss-Ostrogradsky theorem [19] changes the volume integral into an integral over the surface ∂V of volume V and leads to

$$0 = \oint_{\partial V} \mathbf{D} \cdot \mathbf{n} \mathrm{d}a = \left(\mathbf{D}_2 - \mathbf{D}_1\right) \cdot \mathbf{n}\, \Delta a\,, \tag{2.15a}$$

$$0 = \oint_{\partial V} \mathbf{B} \cdot \mathbf{n} \mathrm{d}a = \left(\mathbf{B}_2 - \mathbf{B}_1\right) \cdot \mathbf{n}\, \Delta a\,, \tag{2.15b}$$

where the surface normal vector $\mathbf{n}$ always points outwards and the fields' subscripts indicate the respective medium. In Eqs. (2.15) we already evaluated the surface integral in the limit of $\delta \to 0$ and took into account that the normal vector gives rise to contributions opposite in sign between top and bottom faces. As an immediate consequence, the components normal to the media interface

$$\mathbf{D}_\perp \quad \text{and} \quad \mathbf{B}_\perp \quad \text{are continuous.} \tag{2.16}$$

A similar procedure can be applied to Maxwell's source free curl equations Eq. (2.4c) and Eq. (2.4d), but this time we integrate over an area S perpendicular to the material interface. Using Stokes'

theorem [19] to dispose the curl operator, the integral changes to an integral over the edge ∂S of the area. Again, we pick a particular closed integration contour – the Stokesian loop – depicted in Fig. 2.1. Its top and bottom edges of length Δl are chosen parallel to the material interface, the remaining two edges of length δ normal to it. Then, the resulting equations

$$\oint_{\partial S} \mathbf{E}(\mathbf{r},\omega)\cdot \mathrm{d}\mathbf{l} = \quad \mathrm{i}\omega \int_S \mathbf{B}(\mathbf{r},\omega)\cdot \mathbf{n}\mathrm{d}a\,, \tag{2.17a}$$

$$\oint_{\partial S} \mathbf{H}(\mathbf{r},\omega)\cdot \mathrm{d}\mathbf{l} = -\mathrm{i}\omega \int_S \mathbf{D}(\mathbf{r},\omega)\cdot \mathbf{n}\mathrm{d}a\,, \tag{2.17b}$$

are in the limit of $\delta \to 0$, where the integration area on the right hand side vanishes, evaluated to

$$\big(\mathbf{E}_1 - \mathbf{E}_2\big)\cdot \mathbf{t}\ \Delta l = 0\,, \tag{2.18a}$$

$$\big(\mathbf{H}_1 - \mathbf{H}_2\big)\cdot \mathbf{t}\ \Delta l = 0\,. \tag{2.18b}$$

Again, the left and right edge vanish, and the top and bottom edge contributions feature an opposite sign because of the directed contour which points along the tangential vector $\mathbf{t} = \mathbf{n}_{loop} \times \mathbf{n}$. Hence, the tangential components

$$\mathbf{E}_{\parallel} \qquad \text{and} \qquad \mathbf{H}_{\parallel} \qquad \text{are continuous} \tag{2.19}$$

across the interface as well.

The continuity conditions Eq. (2.16) and Eq. (2.19) are an important result. They will become an essential condition in the derivation of the scattering matrix algorithm presented in Chap. 5.

2.3. Energy Transport

One of the most important concepts in physics is the concept of energy and energy conservation. Since electromagnetic fields carry energy, we are especially interested in a quantity that describes how much energy they transport through space. The conservation law of energy for electromagnetic fields was first described by John H. Poynting in his publication "On the Transfer of Energy in the Electromagnetic Field" in 1884 [25]. In the integral form it reads for time dependent fields in linear dispersionless media

$$\int_V \partial_t u(\mathbf{r},t)\,\mathrm{d}\mathbf{r} \ = -\int_V \mathbf{j}_\mathrm{f}(\mathbf{r},t)\cdot\mathbf{E}(\mathbf{r},t)\,\mathrm{d}\mathbf{r} \ - \oint_{\partial V} \mathbf{S}(\mathbf{r},t)\cdot\mathbf{n}\,\mathrm{dA}\,. \tag{2.20}$$

The neat interpretation is as follows: A change in the energy content of the electromagnetic fields given by time derivative of the energy density

$$u(\mathbf{r},t) = \frac{1}{2}\big(\mathbf{E}(\mathbf{r},t)\cdot\mathbf{D}(\mathbf{r},t) + \mathbf{H}(\mathbf{r},t)\cdot\mathbf{B}(\mathbf{r},t)\big) \tag{2.21}$$

within the finite volume V can either originate from work done by the electric field to free charges and thus conversion into mechanical or thermal energy given by $\mathbf{j}_\mathrm{f}\cdot\mathbf{E}$, or from an energy flux through the surface ∂V of the volume given by the *Poynting vector*

$$\mathbf{S}(\mathbf{r},t) = \mathbf{E}(\mathbf{r},t)\times\mathbf{H}(\mathbf{r},t)\,. \tag{2.22}$$

As stated before, this form of Poynting's theorem and its interpretation is only valid in dispersionless media. But the Poynting vector $\mathbf{S}$ retains its form and interpretation even for dispersive media. Since we are mainly interested in the energy transport properties of time harmonic electromagnetic waves, only the *time-averaged* Poynting vector

$$\mathbf{S}(\mathbf{r},\omega) := \big\langle \mathbf{S}(\mathbf{r},t) \big\rangle_t = \frac{1}{2}\mathrm{Re}\left(\mathbf{E}(\mathbf{r},\omega) \times \mathbf{H}^*(\mathbf{r},\omega)\right) \tag{2.23}$$

is important to us. Here, the asterisk * denotes the complex conjugate, and the time average $\langle\cdot\rangle_t$ is a normalized time integral over one full period $T = 2\pi/\omega$.

2.4. Electromagnetic Waves

Electromagnetic waves are solutions to Maxwell's equations. They consist of coupled electric and magnetic fields that propagate through space. Their behavior is governed by the wave equation derived in Sec. 2.4.1. For homogeneous isotropic media the problem simplifies and the solutions are given by linearly polarized plane waves introduced in Sec. 2.4.2. Section 2.4.3 finally discusses the continuity of plane waves at material interfaces.

2.4.1. Wave Equation

Maxwell's curl equations in the form of Eqs. (2.13) can be combined and recast into the wave equation. The wave equation can be formulated in two versions either depending on the electric field $\mathbf{E}$ or the magnetic field $\mathbf{H}$. Both versions are analytically equivalent but the numerical evaluation of the latter is often less complicated. Here, their derivation is sketched for the $\mathbf{E}$-field only but similarly applies to the $\mathbf{H}$ field analog.

Thereto, we multiply Eq. (2.13a) by the inverse permeability tensor $\underline{\boldsymbol{\mu}}^{-1}$ from the left and take the curl of both sides. Replacing $\boldsymbol{\nabla} \times \mathbf{H}$ on the right hand side with Eq. (2.13b) leads to

$$\boldsymbol{\nabla} \times \left[\underline{\boldsymbol{\mu}}^{-1}(\mathbf{r},\omega) \cdot \left(\boldsymbol{\nabla} \times \mathbf{E}(\mathbf{r},\omega)\right)\right] = \omega^2 \underline{\boldsymbol{\varepsilon}}(\mathbf{r},\omega)\mathbf{E}(\mathbf{r},\omega)\,, \tag{2.24a}$$

$$\boldsymbol{\nabla} \times \left[\underline{\boldsymbol{\varepsilon}}^{-1}(\mathbf{r},\omega) \cdot \left(\boldsymbol{\nabla} \times \mathbf{H}(\mathbf{r},\omega)\right)\right] = \omega^2 \underline{\boldsymbol{\mu}}(\mathbf{r},\omega)\mathbf{H}(\mathbf{r},\omega)\,. \tag{2.24b}$$

Only one of the above equations needs to be solved, the missing field can be calculated using the appropriate relation of Eqs. (2.13). In the subsequent section we will use the wave equation for the $\mathbf{E}$-field.

2.4.2. Plane Waves

In the special case of homogeneous isotropic material properties where permittivity $\underline{\boldsymbol{\varepsilon}}(\mathbf{r},\omega) = \epsilon(\omega)$ and permeability $\underline{\boldsymbol{\mu}}(\mathbf{r},\omega) = \mu(\omega)$ reduce to scalar quantities, the $\mathbf{E}$-field wave equation, Eq. (2.24a) simplifies to

$$\boldsymbol{\nabla} \times \left(\boldsymbol{\nabla} \times \mathbf{E}(\mathbf{r},\omega)\right) = \omega^2\, \epsilon(\omega)\mu(\omega)\, \mathbf{E}(\mathbf{r},\omega)\,. \tag{2.25}$$

A solution to this partial differential equation is given by a plane wave

$$\mathbf{E}(\mathbf{r},\omega) = \mathbf{E}_0(\omega) \cdot e^{\mathrm{i}\mathbf{k}\mathbf{r}}\,, \tag{2.26}$$

featuring a frequency-dependent complex amplitude $\mathbf{E}_0(\omega) = E_0(\omega)\hat{\mathbf{E}}_0$ with polarization unit vector $\hat{\mathbf{E}}_0$, and the complex wave vector $\mathbf{k} = k\,\hat{\mathbf{k}}$ with propagation direction unit vector $\hat{\mathbf{k}}$. The name plane wave derives from the property that surfaces of constant amplitude and phase $\phi = \mathbf{kr}$ form planes which are parallel to each other and normal to $\hat{\mathbf{k}}$. Amplitude and wave vector are both subject to some additional constraints that can be seen when inserting Eq. (2.26) into Eq. (2.25). This, using the vector identity $\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \boldsymbol{\Psi}) = \boldsymbol{\nabla}(\boldsymbol{\nabla} \cdot \boldsymbol{\Psi}) - \Delta\boldsymbol{\Psi}$, entails the condition

$$-k^2\,\hat{\mathbf{k}}\left(\hat{\mathbf{k}}\cdot\hat{\mathbf{E}}_0\right) + k^2\,\left(\hat{\mathbf{k}}\cdot\hat{\mathbf{k}}\right)\hat{\mathbf{E}}_0 \stackrel{!}{=} \omega^2\,\epsilon\mu\,\hat{\mathbf{E}}_0\,. \tag{2.27}$$

Because of the absence of sources, Maxwell's divergence equation, Eq. (2.4a), guarantees $\boldsymbol{\nabla}\cdot\mathbf{E} = 0$ and, thus, as first constraint, we deduce

$$\hat{\mathbf{k}}\cdot\hat{\mathbf{E}}_0 = 0\,, \tag{2.28}$$

from the first term on the left hand side which means that the polarization of the electric field must always be perpendicular to the propagation direction of the wave. Furthermore, the polarization of the magnetic field

$$\mathbf{H} = \frac{k\,E_0}{\mu\,\omega^2}\;\left(\hat{\mathbf{k}}\times\hat{\mathbf{E}}_0\right) \tag{2.29}$$

can be calculated from Eq. (2.13a). Consequently, $\mathbf{E}$, $\mathbf{H}$ and $\mathbf{k}$ must be *mutually orthogonal*, plus electric and magnetic field strength have always a fixed ratio throughout space – we say they are *in phase*.

The second constraint stems from the remainder of Eq. (2.27) which provides the *dispersion relation*

$$k^2 = \omega^2\,\epsilon(\omega)\mu(\omega)\,, \tag{2.30}$$

assuming

$$\hat{\mathbf{k}}\cdot\hat{\mathbf{k}} = 1\,. \tag{2.31}$$

Hence, the length of the wave vector – the wave number k – as well as the wavelength $\lambda = 2\pi/k$ are related to the frequency of the plane wave by the material properties or more precise the speed of light

$$c(\omega) = \frac{1}{\sqrt{\epsilon(\omega)\mu(\omega)}} = \frac{1}{n(\omega)} \tag{2.32}$$

in the medium, or similarly the refractive index $n(\omega)$ which is a property of plane waves. Note, that in dimensionless units the free space frequency equals the free space wave number, whereas in SI units the dispersion relation Eq. (2.30) gains a factor ${c_0}^2$ on the left hand side. Mind as well, that we allowed for dispersive materials which makes the wave number in general complex. Assuming real unit vectors and for simplicity $\mu(\omega) = 1$, the Poynting vector Eq. (2.23) together with Eq. (2.26) and Eq. (2.29) yields

$$\mathbf{S}(\mathbf{r},\omega) = \frac{\left|E_0\right|^2}{2\omega^2}\,\mathrm{Re}(k)\exp\left[-2\,\mathrm{Im}(k)\,r_{\hat{\mathbf{k}}}\right]\,\hat{\mathbf{k}}\,, \tag{2.33}$$

with $r_{\hat{\mathbf{k}}} = \hat{\mathbf{k}}\cdot\mathbf{r}$.

As stated before, the wave vector can be complex-valued which also includes purely real or purely imaginary. When referring to plane waves one usually means *propagating waves* with purely real

$\mathbf{k}$-vectors and oscillating electromagnetic fields of constant amplitude. They are able to transport energy even far away from their source. Waves with purely imaginary $\mathbf{k}$-vector are called *evanescent waves* and are present in the near field region around a source only. They do not transport energy, because their field amplitudes decay exponentially. Merely in the context of media with complex material parameters, there exists a third type of waves whose wave vector has both a real and an imaginary portion. These *damped waves* are oscillatory and transport energy but, because of absorption in the material, field amplitudes and energy current density slowly decay[5].

The above discussion of complex wave vectors emanated from the assumption that the wave number was complex and the direction vector $\hat{\mathbf{k}}$ was real. In principle, Eq. (2.31) is also fulfilled by a properly chosen complex propagation direction unit vector $\hat{\mathbf{k}} = \mathbf{k}_\mathrm{R} + \mathrm{i}\mathbf{k}_\mathrm{I}$. The most fundamental solution to the wave equations, Eqs. (2.24), is then given by the exponential term

$$\begin{aligned}
\exp\left[\mathrm{i}\mathbf{k}\mathbf{r}\right] &= \exp\left[\mathrm{i}(a+\mathrm{i}b)\left(\mathbf{k}_\mathrm{R}+\mathrm{i}\mathbf{k}_\mathrm{I}\right)\mathbf{r}\right] , \\
&= \exp\left[\mathrm{i}\left(a\mathbf{k}_\mathrm{R}-b\mathbf{k}_\mathrm{I}\right)\mathbf{r}\right]\ \exp\left[-\left(b\mathbf{k}_\mathrm{R}+a\mathbf{k}_\mathrm{I}\right)\mathbf{r}\right] , && \text{(2.34a)} \\
&\overset{b=0}{\Rightarrow}\ \exp\left[\mathrm{i}a\mathbf{k}_\mathrm{R}\mathbf{r}\right]\ \exp\left[-a\mathbf{k}_\mathrm{I}\mathbf{r}\right] , && \text{(2.34b)} \\
&\overset{\mathbf{k}_\mathrm{I}=0}{\Rightarrow}\ \exp\left[\mathrm{i}a\mathbf{k}_\mathrm{R}\mathbf{r}\right]\ \exp\left[-b\mathbf{k}_\mathrm{R}\mathbf{r}\right] , && \text{(2.34c)}
\end{aligned}$$

with $k = a + \mathrm{i}b$, and $\mathbf{k}_\mathrm{R} \cdot \mathbf{k}_\mathrm{I} = 0$. This generalization leads to *inhomogeneous plane waves* whose surfaces of constant amplitude and phase still form planes but are not parallel anymore (cf. Eq. (2.34a) and Eq. (2.34b)). However, inhomogeneous plane wave solutions are not topic of this thesis. The interested reader may refer to Ref. [26] for detailed information instead.

2.4.3. Continuity at Material Interfaces

In Sec. 2.2 we derived the continuity conditions for electromagnetic fields at material interfaces. These general consideration equally apply to electromagnetic waves. In this section we contemplate the implications that follow for a plane wave of frequency ω and wave vector $\mathbf{k}$ incident from region $\mathbf{1}$ ($z < 0$) with isotropic linear material properties condensed in the refractive index $n_\mathbf{1}(\omega) = \sqrt{\varepsilon_\mathbf{1}(\omega)\,\mu_\mathbf{1}(\omega)}$ onto an interface with region $\mathbf{2}$ ($z > 0$) with $n_\mathbf{2}(\omega) = \sqrt{\varepsilon_\mathbf{2}(\omega)\,\mu_\mathbf{2}(\omega)}$ correspondingly. For simplicity the interface is assumed to be flat and positioned at $z = 0$ with interface normal vector $\mathbf{n}$.

The physical observation is that one part of the incident plane wave is reflected from the interface and travels back with wave vector $\mathbf{k}''$. A second part is transmitted into region $\mathbf{2}$ traveling there with wave vector $\mathbf{k}'$. The situation is depicted in Fig. 2.2.

Continuity of Wave Vectors

Let us consider the implications of the field continuity conditions Eq. (2.16) and Eq. (2.19) for these wave vectors. These conditions must be fulfilled everywhere on the interface at all times which

[5] Field amplitudes and energy content could also increase in certain materials. These materials which act as electromagnetic field sources by converting energy from other forms like electrical energy are called gain or active materials. We do not treat them in the course of this thesis. The materials we treat, transparent or absorptive, are called passive materials.

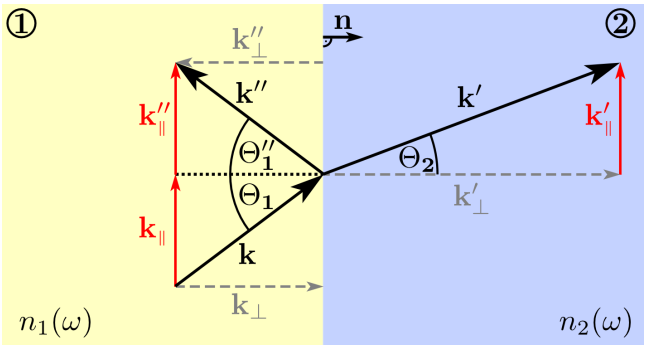


Figure 2.2.: *Continuity of* $\mathbf{k}$*-vectors at a material interface.*

means that the spatial variation of all fields in the immediate vicinity of the interface in both regions must be equal. Because the spatial distribution of plane waves is solely governed by the exponential term, the phases of all three waves must consequently be equal

$$(\mathbf{k} \cdot \mathbf{r})_{z=0} = (\mathbf{k}' \cdot \mathbf{r})_{z=0} = (\mathbf{k}'' \cdot \mathbf{r})_{z=0} \,. \tag{2.35}$$

On the one hand, this implies that all wave vectors must lie in one plane, the plane of incidence, and on the other hand, we deduce that their components *parallel* to the interface must be conserved.

Since $|\mathbf{k}''| = |\mathbf{k}|$, the latter provides that the angle of incidence equals the angle of reflection

$$\theta_\mathbf{1} = \theta_\mathbf{1}'' \,, \tag{2.36}$$

where $\theta_\mathbf{1}$ and $\theta_\mathbf{1}''$ are defined as the angles the wave vectors incur with the interface normal $\mathbf{n}$. And secondly, we obtain Snell's law of refraction

$$\frac{\sin\theta_\mathbf{1}}{\sin\theta_\mathbf{2}} = \frac{n_\mathbf{2}}{n_\mathbf{1}} = \frac{c_\mathbf{1}}{c_\mathbf{2}} \,, \tag{2.37}$$

which describes the *refraction* of plane waves – the change of propagation direction according to the ratio between the speeds of light because of differing material properties – at material interfaces.

Total Internal Reflection

Let us consider the situation when the refractive index in region $\mathbf{1}$ is larger than in region $\mathbf{2}$. Then $n_\mathbf{1} > n_\mathbf{2}$ and $\theta_\mathbf{2} > \theta_\mathbf{1}$ according to Snell's law Eq. (2.37). Consequently, there exists an angle $\theta_\mathbf{1} = \theta_\mathrm{TIR}$ with

$$\theta_\mathrm{TIR} = \arcsin\left(\frac{n_\mathbf{2}}{n_\mathbf{1}}\right) \tag{2.38}$$

when $\theta_\mathbf{2} = \pi/2$, which means that the refracted wave travels parallel to the interface and no energy is transported across the interface. Thus, for that angle there must be *total internal reflection*, where internal refers to the medium with the higher refractive index.

Looking at Eq. (2.37), for angles $\theta_\mathbf{1} > \theta_\mathrm{TIR}$ we note that $\sin\theta_\mathbf{2}$ must be larger than one. From the identity relation $(\sin^2\theta_\mathbf{2} + \cos^2\theta_\mathbf{2}) = 1$ we deduce that, in this situation, the cosine must be purely

imaginary. In a similar fashion as we derived Eq. (2.33) the Poynting flux across the interface with normal vector $\mathbf{n}$ is obtained as

$$\mathbf{n} \cdot \mathbf{S}(\mathbf{r}, \omega) \propto \mathrm{Re}[\underbrace{\mathbf{n} \cdot \mathbf{k}'}_{k' \cos \theta_\mathbf{2}}] = 0\,, \tag{2.39}$$

and evaluates to zero because of the purely imaginary cosine. The electromagnetic fields in region **2** decay exponentially in the direction normal to the interface as can be seen by examining the exponential term

$$e^{\mathrm{i}\mathbf{k}'\mathbf{r}} = e^{\mathrm{i}k'(x \sin\theta_\mathbf{2} + z\cos\theta_\mathbf{2})} = e^{\mathrm{i}k' \sin\theta_\mathbf{2}\, x}\; e^{-k'|\cos\theta_\mathbf{2}|\, z}\,. \tag{2.40}$$

Summarizing the above findings, total internal reflection occurs for angles $\theta_\mathbf{1} \geq \theta_{\mathrm{TIR}}$. This concept will become important when we discuss the properties of optical waveguides in Chap. 4.

Fresnel Equations and Polarizations

From the continuity of the fields at the interface we can also derive ratios between amplitudes of transmitted (reflected) fields and the incident field. The corresponding equations are called the *Fresnell equations*. Their derivation and definition is omitted here, because they are not necessary for the discussion of our topic. The interested reader may consult any standard optics textbook as for example Ref. [27].

In the course of their derivation, however, it is convenient to introduce two special cases which we would like to establish here as well. The arbitrarily (linear) polarized plane wave as, for instance, defined by the polarization vector of the electric field $\hat{\mathbf{E}}_0$ with wave vector $\mathbf{k}$ incident onto the interface can be decomposed into two polarizations which are easier to handle. They are sketched in Fig. 2.3. In the *s-polarization* the polarization vector of the electric field is perpendicular to the

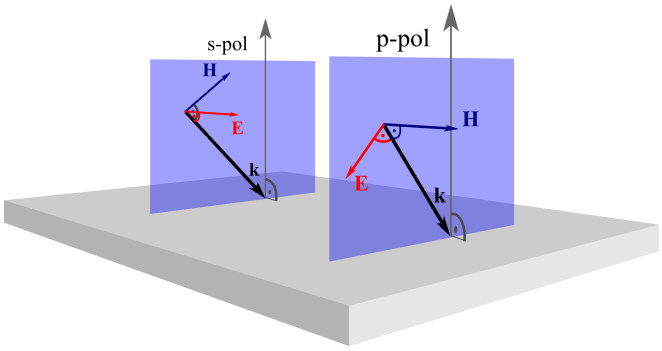


Figure 2.3.: *In s-polarization the electric field is orthogonal to the plane of incidence, and in p-polarization it is parallel to the plane of incidence.*

plane of incidence and the $\mathbf{H}$-field polarization vector is in the plane of incidence. In *p-polarization* both polarization vectors are exactly the other way round, the polarization of the magnetic field is perpendicular to the plane of incidence and the $\mathbf{E}$-field polarization vector is in the plane. This convention and the corresponding decomposition of waves will be used in the discussion about the simulation of periodic systems in Chap. 6.

2.5. Covariant Formulation of Maxwell's Equations

The physical phenomena that are described by Maxwell's equations are independent of the coordinate system used to describe them. A covariant formulation of the equations guarantees that not only the physical properties, but also the formulas have the same appearance in every coordinate system. In order to obtain this appearance, we have to study how the building blocks of Maxwell's equations — scalars, vectors, derivatives, and operators — behave when we change from one coordinate system to another.

The topic of tensor analysis in curvilinear coordinate systems is covered in textbooks, e.g., Ref. [28]. The more specialized related topic of covariant formulation of Maxwell's equations is dealt with in a few textbooks only, like for example Ref. [29]. For a detailed introduction we particularly recommend Ref. [22] because it develops the topic from scratch in small steps covering both aspects. Furthermore, it illustrates the close connection between curvilinear coordinate transformations and transformation optics, which allows for the construction of intriguing applications like "almost" invisibility devices. [30, 31].

Before we start with the presentation, we recall a few important conventions which are used in the following. The *Einstein summation convention* implies a summation over repeated indices, e.g.

$$A^{\rho\sigma}\, B_\sigma \equiv \sum_{\sigma=1}^{3} A^{\rho\sigma}\, B_\sigma = A^{\rho 1} B_1 + A^{\rho 2} B_2 + A^{\rho 3} B_3 \,. \tag{2.41}$$

In the context of this topic it is always assumed that this kind of summation is carried out between one superscript and one subscript index which is then called a *contraction*.

The resulting quantity $C^\rho = A^{\rho\sigma}\, B_\sigma$ has only one remaining index that is *not* summed over – referred to as *free index*. The *Einstein range convention* implies that a free index, like for example index ρ in the last equations, is concerted to bestride all its possible values

$$C^\rho \equiv \{C^\rho,\ \rho = 1, 2, 3\} \,. \tag{2.42}$$

This does not only hold for single quantities but also for whole equation and thus describes sets of equations, e.g. for the components of a vector.

Differently but similarly denoted, function arguments with free indices stand for

$$f(x^\rho) \equiv f(x^1, x^2, x^3) \,. \tag{2.43}$$

The above introduced conventions never apply to those indices directly.

2.5.1. Curvilinear Coordinate Systems

The absolute position in three dimensional space is given by coordinates $\{x^\rho,\ \rho = 1, 2, 3\}$ in any arbitrary coordinate system spanned by its basis vectors $\{\mathbf{e}_\rho,\ \rho = 1, 2, 3\}$ relative to a fixed origin $\mathcal{O}$. Then, we know that the position relative to the origin $\mathcal{O}$ can be described by the position vector $\mathbf{r}$ which is given as

$$\mathbf{r} = x^\rho\, \mathbf{e}_\rho \,. \tag{2.44}$$

Please note that the basis vectors themselves are in general – in curvilinear coordinate systems – not independent of the coordinates $\mathbf{e}_\rho = \mathbf{e}_\rho(x^\sigma)$, and are neither mutually orthogonal nor normalized. These properties are characteristics of Cartesian coordinate systems only.

The length of every vector $|\mathbf{V}|$ is a scalar whose square is given by the scalar product of the vector with itself

$$|\mathbf{V}|^2 = \mathbf{V} \cdot \mathbf{V} = V^\rho V^\sigma \, (\mathbf{e}_\rho \cdot \mathbf{e}_\sigma) \mathrel{\hat{=}} V^\rho V^\sigma \, g_{\rho\sigma} \, . \tag{2.45}$$

Here, we would like to make the reader aware of some peculiarities in the used notations. The term on the left hand side of the "$\hat{=}$" sign, written in the *basis vector notation* of Eq. (2.44), and the term on the right hand side, written in *tensor notation*, look trivially equivalent. In fact, the expressions are equivalent, but the notations have dramatically changed from left to right. A vector in the basis vector notation is given by the weighted sum of the basis vectors. The weights are *scalars* and are *labeled* with a label ρ in order to highlight their affiliation with the corresponding basis vector. This labeling is for convenience and to be able to use the sum convention.

However, in the tensor notation we do not use labels but *indices*, where the number of (free) indices indicates the *rank* of the tensor. The right hand side of the equation is, thus, literally $V^\rho \otimes V^\sigma \otimes g_{\rho\sigma}$ a tensor product between two first rank tensors (vectors) and a second rank tensor (matrix). In this particular case, we have contractable repeated indices which is the characteristic of a generalized scalar product (or dot product) that reduces the rank of the tensor. The rank of the resulting tensor is the number of free indices, which is in the above case zero, and consequently the result is a scalar like indicated on the left hand side by the scalar product symbol "$\cdot$".

The scalar product of the basis vectors $\mathbf{e}_\rho$ and $\mathbf{e}_\sigma$ defines the *metric tensor* element $g_{\rho\sigma}$ which expresses the measure of length in an arbitrary coordinate system. It accounts for the effect that the basis vectors are not normalized, that they are not mutually orthogonal, and that the basis vectors change direction independently with position. Thus, the metric tensor $g_{\rho\sigma} = g_{\rho\sigma}(x^\tau)$ is position dependent itself. It turns out it is also symmetric.

To be consistent with the conventions established above, in order to obtain a scalar, indices in Eq. (2.45) have to be contracted. We deduce that the metric tensor can thus be used to lower index positions as

$$V_\rho = g_{\rho\sigma} V^\sigma \, , \tag{2.46}$$

and the scalar product of Eq. (2.45) can be written like

$$|\mathbf{V}|^2 = V_\rho V^\rho = V^\rho V_\rho \, . \tag{2.47}$$

With the help of the inverse metric tensor $g^{\rho\sigma}$ defined by

$$g^{\rho\sigma} g_{\sigma\tau} = \delta^\rho_\tau \, , \tag{2.48}$$

where δ^ρ_σ is the *scalar* Kronecker delta with value one if $\rho = \sigma$ and zero else, it is similarly possible to raise index positions. The lower index quantities V_ρ are the components of a vector constituted from basis vectors with upper indices. For each individual coordinate system, its *dual vector space* with dual basis vectors $\mathbf{e}^\rho = \mathbf{e}^\rho(x_\rho)$ can be defined by

$$\mathbf{e}^\rho \equiv \frac{1}{2} \, \frac{\epsilon^{\rho\sigma\tau}}{\det \underline{\mathbf{J}}} \, \mathbf{e}_\sigma \times \mathbf{e}_\tau \, , \tag{2.49}$$

where we introduced the permutation symbol

$$\epsilon_{\rho\sigma\tau} = \epsilon^{\rho\sigma\tau} \equiv \begin{cases} +1 & \text{for} \quad \rho\sigma\tau \text{ is an even permutation of } 123\,, \\ -1 & \text{for} \quad \rho\sigma\tau \text{ is an odd permutation of } 123\,, \\ 0 & \text{else}\,, \end{cases} \tag{2.50}$$

and the Jacobian matrix $\underline{\mathbf{J}}$ whose determinant provides the volume element spanned by the basis vectors $\mathbf{e}_\rho$ and is alternatively expressed through their triple product $\det \underline{\mathbf{J}} = 1/2\, \epsilon^{\rho\sigma\tau}\, \mathbf{e}_\rho \cdot (\mathbf{e}_\sigma \times \mathbf{e}_\tau)$. The summation convention is applied in Eq. (2.49). The permutation symbol simply allows us to write the expressions

$$\mathbf{e}^1 \equiv \frac{\mathbf{e}_2 \times \mathbf{e}_3}{\det \underline{\mathbf{J}}} = -\frac{\mathbf{e}_3 \times \mathbf{e}_2}{\det \underline{\mathbf{J}}} \tag{2.51}$$

and all similar expressions with cyclic permutations of the labels in a short form.

The definition of the dual basis in Eq. (2.49) fulfills the orthogonality relation

$$\mathbf{e}^\rho \cdot \mathbf{e}_\sigma = \mathbf{e}_\sigma \cdot \mathbf{e}^\rho = \delta^\rho_\sigma\,. \tag{2.52}$$

The basis vectors and their orthogonality are schematically depicted in Fig. 2.4. From the rewritten

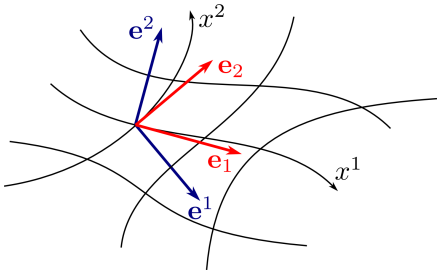


Figure 2.4.: *Sketch of the basis vectors $\mathbf{e}_\rho$ and the dual basis vectors $\mathbf{e}^\rho$. Basis vectors are tangential to the coordinate lines, whereas the dual basis vectors fulfill the orthogonality relation Eq. (2.52). The direction of the basis vectors is position dependent.*

Eq. (2.47) with respect to the orthogonality relation Eq. (2.52), namely

$$|\mathbf{V}|^2 = (V_\rho\, \mathbf{e}^\rho) \cdot (V^\sigma\, \mathbf{e}_\sigma)\,, \tag{2.53}$$

we catch the most important property of the whole notational scheme: the dual basis and its vector components are defined such that, as long as we contract only upper and lower index components, the lengths and distances are conserved and the metric tensor is properly accounted for, independent of position and chosen coordinate system – like we are used to from Cartesian coordinate systems.

Later on we will often encounter vectors as dot products of higher rank tensors. Therefore we make an easy example that sketches the general procedure for those products. In the basis vector notation, the equation $\mathbf{C} = \underline{\mathbf{A}} \cdot \mathbf{B}$ is consistently written as

$$\begin{aligned} C^\rho\, \mathbf{e}_\rho &= (A^{\rho\sigma}\, \mathbf{e}_\rho \otimes \mathbf{e}_\sigma) \cdot (B_\tau\, \mathbf{e}^\tau) \\ &= A^{\rho\sigma}\, B_\tau\, \mathbf{e}_\rho \otimes (\mathbf{e}_\sigma \cdot \mathbf{e}^\tau) \\ &= A^{\rho\sigma}\, B_\tau\, \mathbf{e}_\rho \otimes \delta^\sigma_\tau \\ &= A^{\rho\sigma}\, B_\sigma\, \mathbf{e}_\rho\,, \end{aligned} \tag{2.54}$$

which in turn in the tensor notation is equivalent to

$$C^{\rho} = A^{\rho\sigma}\, B_{\sigma}\,. \tag{2.55}$$

It is legitimate to conclude that it is possible to change between both notations easily, as long as we are aware of the different meaning of labels and indices. We stress once again, that most of the products we encounter in the equations are generalized scalar products between a quantity and its dual counterpart. This approach leaves the formulated equation invariant of the chosen coordinate system.

Before we can proceed with the issue of how tensor quantities of different rank change, subject to a coordinate transformation, there is one important piece missing which we have to introduce first: derivatives. The discussion is started with a scalar field that is a function of the coordinates. We can take its derivative

$$\frac{\partial}{\partial x^{\rho}}\,\psi(x^{\rho}) \equiv \partial_{\rho}\,\psi(x^{\rho}) \tag{2.56}$$

with respect to x^{ρ} and attribute this operation a new operator ∂_{ρ} in tensor notation. For reasons that become clear later, the new operator is a rank one tensor with one *subscript* index and hence a quantity from the dual vector space – the ρ-th component of the dual gradient. Then the *gradient* in basis vector notation using the inverse metric tensor to raise the index reads

$$\boldsymbol{\nabla}\psi = g^{\rho\sigma}\,(\partial_{\sigma}\,\psi(x^{\rho}))\;\mathbf{e}_{\rho}\,. \tag{2.57}$$

For the derivative of a vector field $\mathbf{V}(x^{\rho})$ we have to keep in mind that both the scalar field describing the vector component as well as the basis vectors depend on the coordinates. The differentiation must consequently obey the product rule

$$\frac{\partial}{\partial x^{\rho}}\mathbf{V}(x^{\rho}) = \left(\frac{\partial}{\partial x^{\rho}}\,V^{\sigma}(x^{\rho})\right)\mathbf{e}_{\sigma}(x^{\rho}) + V^{\sigma}(x^{\rho})\left(\frac{\partial}{\partial x^{\rho}}\,\mathbf{e}_{\sigma}(x^{\rho})\right). \tag{2.58}$$

Because the derivative of a vector is a vector itself, the part in brackets of the rightmost term can be written in basis vector notation as

$$\frac{\partial}{\partial x^{\rho}}\,\mathbf{e}_{\sigma} = \Gamma^{\tau}_{\sigma\rho}\,\mathbf{e}_{\tau}\,, \tag{2.59}$$

where the quantity $\Gamma^{\tau}_{\sigma\rho}$ is the τ-th component of the derivative of $\mathbf{e}_{\sigma}$ with respect to x^{ρ} called *Christoffel symbol*. They are symmetric in the lower labels $\Gamma^{\tau}_{\sigma\rho} = \Gamma^{\tau}_{\rho\sigma}$, and can be calculated from the metric tensor via the relation

$$\Gamma^{\sigma}_{\tau\rho} = \frac{1}{2}g^{\sigma\kappa}(\partial_{\rho}\,g_{\kappa\tau} + \partial_{\tau}\,g_{\kappa\rho} - \partial_{\kappa}\,g_{\tau\rho})\,. \tag{2.60}$$

The 27 Christoffel symbols can be used to rewrite Eq. (2.58) into

$$\frac{\partial}{\partial x^{\rho}}\mathbf{V}(x^{\rho}) = \left(\partial_{\rho}\,V^{\sigma} + \Gamma^{\sigma}_{\tau\rho}V^{\tau}\right)\mathbf{e}_{\sigma} \equiv (\nabla_{\rho}\,V^{\sigma})\,\mathbf{e}_{\sigma}\,, \tag{2.61}$$

where we defined the *covariant derivative operator* ∇_{ρ}. We see that the covariant derivative is the right way to differentiate a vector since it considers the spatial dependence of both, the vector component and the basis vector. The covariant derivative of the dual vector is similarly defined as

$$\nabla_{\rho}\,U_{\sigma} \equiv \partial_{\rho}\,U_{\sigma} - \Gamma^{\tau}_{\sigma\rho}U_{\tau}\,. \tag{2.62}$$

We would like to note that the Christoffel symbols themselves are not a tensor quantity even though we use the sum convention in the equations. Only their sums and differences in the covariant derivatives have tensor properties. This means $\nabla_\rho V_\sigma$ which is sometime alternatively denoted as $V_{\sigma;\rho}$ constitutes a rank two tensor with all the accompanying properties described above, in particular the raising and lowering of indices with the help of the metric tensor. It needs to be stressed that the covariant derivative's form crucially depends on the tensor it is applied to, whether it is a tensor or a dual tensor and its rank.

Higher order tensors are differentiated according to the following rule: subscript indices get a minus sign in front of the Christoffel symbol like in Eq. (2.62), superscript indices get a positive sign in front of the Christoffel symbol as in Eq. (2.61). Thus, a second rank mixed tensor $A_\rho{}^\sigma$ is differentiated like

$$\nabla_\tau A_\rho{}^\sigma = \partial_\tau A_\rho{}^\sigma - \Gamma^\kappa_{\rho\tau} A_\kappa{}^\sigma + \Gamma^\sigma_{\kappa\tau} A_\rho{}^\kappa , \tag{2.63}$$

and higher rank tensors accordingly. The covariant derivatives of the metric tensor vanish

$$\nabla_\tau g_{\rho\sigma} = \nabla_\tau g^{\rho\sigma} = 0 , \tag{2.64}$$

independent of the chosen coordinate system.

With the covariant derivative and these rules at hand, it is finally straight forward to define the *divergence* of a vector $\mathbf{V}$ in covariant notation. It follows from Eq. (2.61) for contracted indices such as

$$\boldsymbol{\nabla} \cdot \mathbf{V} = \nabla_\rho V^\rho = \partial_\rho V^\rho + \Gamma^\rho_{\sigma\rho} V^\sigma = \frac{1}{\sqrt{g}} \partial_\rho (\sqrt{g} V^\rho) , \tag{2.65}$$

where the rightmost term is its simplified but nonetheless convenient version.

The *vector product* between two vectors $\mathbf{U}$ and $\mathbf{V}$ in arbitrary curvilinear coordinates is obtained from

$$\mathbf{U} \times \mathbf{V} = U_\sigma V_\tau \, (\mathbf{e}^\sigma \times \mathbf{e}^\tau) = \xi^{\rho\sigma\tau} U_\sigma V_\tau \, \mathbf{e}_\rho , \tag{2.66}$$

where we used for the last step the dual analog of Eq. (2.49).[6] Furthermore, we introduced the *Levi-Civita tensor*

$$\xi^{\rho\sigma\tau} = \epsilon^{\rho\sigma\tau} / \det \underline{\mathbf{J}} , \qquad \xi_{\rho\sigma\tau} = \det \underline{\mathbf{J}} \, \epsilon^{\rho\sigma\tau} , \tag{2.67}$$

as generalization to arbitrary curvilinear coordinate systems of the permutation operator which determines the vector product in Cartesian coordinate systems.

With this, we can just as well note down the *curl* of a vector in basis vector notation as

$$(\boldsymbol{\nabla} \times \mathbf{V})^\rho = \xi^{\rho\sigma\tau} \nabla_\sigma V_\tau = \xi^{\rho\sigma\tau} \partial_\sigma V_\tau , \tag{2.68}$$

where the covariant derivative luckily simplifies to a derivative of the vector component because the Christoffel symbol terms cancel each other.

At last, we have all the building blocks at hand to rewrite Maxwell's equations in the covariant form in Sec. 2.5.3. But before we get there we would like to show how the introduced quantities change under coordinate transformations in the subsequent section.

[6] To obtain this, we vector multiply Eq. (2.49) from the right with the basis vector $\mathbf{e}^\sigma$, use the identity $(\mathbf{B} \times \mathbf{C}) \times \mathbf{A} = \mathbf{C}(\mathbf{A} \cdot \mathbf{B}) - \mathbf{B}(\mathbf{A} \cdot \mathbf{C})$, and the orthogonality relations Eq. (2.52). Finally, the indices are relabeled.

2.5.2. Coordinate Transformations

There are many situations in physics where it is easy to specify a problem in a coordinate system, often the Cartesian coordinate system because we have an intuitive understanding of it, but its solution is difficult and requires a lot of effort. Then it is often helpful to solve the problem in a coordinate system specifically adapted to the requirements of the problem. If we have such a suitable coordinate system the only task is to define the problem with respect to it. In most cases, the easiest way to do this is to specify the quantities which describe the problem in the original frame and perform a *coordinate transformation* into the suitable coordinate system. The task ahead is to show how the quantities change under such a coordinate transformation.

Therefore, we introduce a second arbitrary curvilinear coordinate system besides the already in Sec. 2.5.1 introduced arbitrary curvilinear coordinate system $\mathcal{O}x^1x^2x^3$. If we leave the origin $\mathcal{O}$ unchanged, the second coordinate system can be represented by basis vectors $\{\bar{\mathbf{e}}_{\rho'},\, \rho' = 1,2,3\}$ with coordinates $\{\bar{x}^{\rho'},\, \rho' = 1,2,3\}$. We establish the convention that the latter coordinate system denotes the original and the former the problem adapted coordinate system. The coordinate transformation between $\mathcal{O}\bar{x}^1\bar{x}^2\bar{x}^3$ and $\mathcal{O}x^1x^2x^3$ is then given by a set of functions

$$\bar{x}^{\rho} = \bar{x}^{\rho}(x^{\rho})\,, \tag{2.69}$$

or their (locally) inverse functions

$$x^{\rho} = x^{\rho}(\bar{x}^{\rho})\,, \tag{2.70}$$

which unambiguously map the coordinates of both coordinate systems.

The simplest transformation properties can be observed if we use the chain rule to write the connection between the differentials

$$\mathrm{d}x^{\rho} = \frac{\partial x^{\rho}}{\partial \bar{x}^{\rho'}}\,\mathrm{d}\bar{x}^{\rho'}\,, \qquad\qquad \mathrm{d}\bar{x}^{\rho'} = \frac{\partial \bar{x}^{\rho'}}{\partial x^{\rho}}\,\mathrm{d}x^{\rho}\,, \tag{2.71}$$

of the two coordinate sets x^{ρ} and $\bar{x}^{\rho'}$ and the corresponding differential operators

$$\frac{\partial}{\partial x^{\rho}} = \frac{\partial \bar{x}^{\rho'}}{\partial x^{\rho}}\,\frac{\partial}{\partial \bar{x}^{\rho'}}\,, \qquad\qquad \frac{\partial}{\partial \bar{x}^{\rho'}} = \frac{\partial x^{\rho}}{\partial \bar{x}^{\rho'}}\,\frac{\partial}{\partial x^{\rho}}\,. \tag{2.72}$$

Here, we introduce the transformation matrices in Eq. (2.71) and Eq. (2.72) denoted as

$$\Lambda^{\rho'}{}_{\rho} = \frac{\partial \bar{x}^{\rho'}}{\partial x^{\rho}}\,, \qquad\qquad \bar{\Lambda}^{\rho}{}_{\rho'} = \frac{\partial x^{\rho}}{\partial \bar{x}^{\rho'}}\,, \tag{2.73}$$

which equal the familiar Jacobian matrix J and its inverse J^{-1}, respectively. Consequently, their product

$$\Lambda^{\rho'}{}_{\rho}\,\bar{\Lambda}^{\rho}{}_{\sigma'} = \delta^{\rho'}_{\sigma'}\,, \qquad\qquad \bar{\Lambda}^{\rho}{}_{\rho'}\,\Lambda^{\rho'}{}_{\sigma} = \delta^{\rho}_{\sigma}\,, \tag{2.74}$$

results in the Kroneker delta, the entries of the unit matrix, and please note that Λ and $\bar{\Lambda}$ are different matrices.

We see that the differential operator $\partial_{\rho} = \Lambda^{\rho'}{}_{\rho}\bar{\partial}_{\rho'}$ as *covariant* tensor of rank one denoted by a subscript index transforms with the Jacobian, whereas a vector

$$V^{\rho} = \bar{\Lambda}^{\rho}{}_{\rho'}\,\bar{V}^{\rho'} \tag{2.75}$$

transforms with its inverse, similar to the differential in Eq. (2.71). Tensors with superscript indices that transform with $\bar{\Lambda}^{\rho}{}_{\rho'}$ are called *contravariant* tensors. This scheme can be generalized to tensors of rank two like for example the covariant metric tensor

$$g_{\rho\sigma} = \Lambda^{\rho'}{}_{\rho}\, \Lambda^{\sigma'}{}_{\sigma}\, \bar{g}_{\rho'\sigma'}\,, \tag{2.76}$$

which transforms with a Jacobian for each subscript index, or even higher rank tensors as for example a four-index mixed tensor

$$D^{\rho}{}_{\sigma\tau\kappa} = \bar{\Lambda}^{\rho}{}_{\rho'}\, \Lambda^{\sigma'}{}_{\sigma}\, \Lambda^{\tau'}{}_{\tau}\, \Lambda^{\kappa'}{}_{\kappa}\, \bar{D}^{\rho'}_{\sigma'\tau'\kappa'} \tag{2.77}$$

that also transforms in accordance with the simple rules established above. These transformation properties actually define not only co- and contravariant, but tensors in general. Even though the Christoffel symbols have three labels, their transformation properties do not obey the simple rules and it is therefore that they constitute no tensors.

Eq. (2.76) rewritten in matrix notation reads

$$\underline{\mathbf{G}} = \underline{\mathbf{\Lambda}}^{T} \bar{\underline{\mathbf{G}}} \Lambda\,. \tag{2.78}$$

If we take the determinante of both sides and denote the coordinate transformed $\det \underline{\mathbf{G}}$ with g, we obtain

$$g = (\det \underline{\mathbf{\Lambda}})^2\, \bar{g} \qquad \Rightarrow \qquad \det \underline{\mathbf{\Lambda}} = \det \underline{\mathbf{J}} = \pm \frac{\sqrt{g}}{\sqrt{\bar{g}}}\,. \tag{2.79}$$

Consequently, since the Levi-Civita tensor in Cartesian coordinates is given by the permutation symbol where $\sqrt{\bar{g}} = 1$, the Levi-Civita tensor in arbitrary coordinate systems, Eq. (2.67), can be expressed with the help of the metric determinant instead of the Jacobian determinant as

$$\xi^{\rho\sigma\tau} = \frac{\epsilon^{\rho\sigma\tau}}{\pm\sqrt{g}}\,. \tag{2.80}$$

The plus or minus sign in front of the square root indicates a change of handedness in the coordinate transformation. Our transformations will always occur between right-handed coordinate systems only – the minus sign is therefore dropped in the remainder.

For a moment, let us switch to the basis vector notation. Then with Eq. (2.44) and Eq. (2.75) from

$$\mathbf{V} = V^{\rho}\mathbf{e}_{\rho} = \bar{V}^{\rho'}\, \bar{\Lambda}^{\rho}{}_{\rho'}\, \mathbf{e}_{\rho} \overset{!}{=} \bar{V}^{\rho'} \bar{\mathbf{e}}_{\rho'} \qquad \Rightarrow \qquad \mathbf{e}_{\rho} = \Lambda^{\rho'}{}_{\rho} \bar{\mathbf{e}}_{\rho'}\,, \tag{2.81}$$

it becomes evident that the basis vectors as lower index quantities (and first rank tensors) transform covariant as expected.

The last part of this section is concerned with the transformation properties of the covariant derivative and derived expressions like divergence and curl. It is needless to say that the covariant derivative, as covariant quantity, transforms accordingly. The divergence of a vector field is a scalar field and must thus be invariant of the chosen coordinate system. This is conveniently respected through the contraction of a co- and contravariant quantity with their inverse transformation properties as can be explicitly seen when transforming Eq. (2.65)

$$\boldsymbol{\nabla} \cdot \mathbf{V} = \nabla_{\rho} V^{\rho} = \underbrace{\Lambda^{\rho'}{}_{\rho} \bar{\Lambda}^{\rho}{}_{\rho'}}_{1}\, \bar{\nabla}_{\rho'}\, \bar{V}^{\rho'} = \underbrace{\Lambda^{\rho'}{}_{\rho} \bar{\Lambda}^{\rho}{}_{\rho'}}_{1}\, \frac{1}{\sqrt{g}}\, \bar{\partial}_{\rho'}(\sqrt{g} \bar{V}^{\rho'}) = \bar{\nabla} \cdot \bar{\mathbf{V}}\,. \tag{2.82}$$

The same applies to the curl of a vector field, Eq. (2.68), considering that the Levi-Civita tensor is of rank three and transforms contravariant. Then the ρ-th component of the curl

$$\begin{aligned}(\boldsymbol{\nabla} \times \mathbf{V})^{\rho} &= \xi^{\rho\sigma\tau}\, \nabla_{\sigma} V_{\tau} \\ &= (\bar{\Lambda}^{\rho}_{\ \rho'} \bar{\Lambda}^{\sigma}_{\ \sigma'} \bar{\Lambda}^{\tau}_{\ \tau'}\, \bar{\epsilon}^{\rho'\sigma'\tau'})\, (\Lambda^{\sigma'}_{\ \sigma}\, \bar{\nabla}_{\sigma'})\, (\Lambda^{\sigma'}_{\ \sigma}\, \bar{V}_{\tau'}) \\ &= \bar{\Lambda}^{\rho}_{\ \rho'}\, \bar{\epsilon}^{\rho'\sigma'\tau'}\, \bar{\nabla}_{\sigma'}\, \bar{V}_{\tau'} \\ &= \bar{\Lambda}^{\rho}_{\ \rho'}\, (\bar{\nabla} \times \bar{\mathbf{V}})^{\rho'} \end{aligned} \tag{2.83}$$

transforms as expected from a contravariant vector component.

With this last transformation rule figured out, we are now ready to get back to our physics problems and formulate Maxwell's equations in covariant notation in the upcoming section.

2.5.3. Maxwell's Equations in Arbitrary Coordinate Systems

After the introduction of the mathematical formalism of differential geometry with the notation of co- and contravariant tensor quantities in Sec. 2.5.1, and the exploration of their behavior under coordinate transformations between two arbitrary curvilinear coordinate systems in Sec. 2.5.2, we have finally everything ready that is necessary to formulate Maxwell's equations in a form independent of the chosen coordinate system.

Maxwell's curl equations, Eqs. (2.13), for time-harmonic fields in source free linear media, in dimensionless units and in covariant tensor notation then read

$$\epsilon^{\rho\sigma\tau}\, \partial_{\sigma}\, E_{\tau}(\mathbf{r},\omega) = \quad \mathrm{i}\omega^2\, \mu^{\rho\sigma}(\mathbf{r},\omega)\, H_{\sigma}(\mathbf{r},\omega)\,, \tag{2.84a}$$

$$\epsilon^{\rho\sigma\tau}\, \partial_{\sigma}\, H_{\tau}(\mathbf{r},\omega) = -\mathrm{i} \quad \varepsilon^{\rho\sigma}(\mathbf{r},\omega)\, E_{\sigma}(\mathbf{r},\omega)\,. \tag{2.84b}$$

It is straight forward to show that this form is independent of the coordinate system. For our purposes it is convenient to absorb the spatially dependent square root of the metric determinant $\sqrt{g(\mathbf{r})}$ which is included in the Levi-Civita tensor into the material terms that are spatially dependent as well. The Levi-Civita tensor then reduces to the spatially constant permutation tensor $\epsilon^{\rho\sigma\tau}$.

Even though we do not need the divergence equations for our numerical purposes, we state them for completeness:

$$\partial_{\rho}\Big(\varepsilon^{\rho\sigma}(\mathbf{r},\omega)\, E_{\sigma}(\mathbf{r},\omega)\Big) = 0\,, \tag{2.85a}$$

$$\partial_{\rho}\Big(\mu^{\rho\sigma}(\mathbf{r},\omega)\, H_{\sigma}(\mathbf{r},\omega)\Big) = 0\,. \tag{2.85b}$$

Furthermore, we also have the tools to find expressions for the material parameters in any adapted coordinate system. Suppose we have the permittivity $\bar{\varepsilon}^{\rho'\sigma'}(\bar{x}^{\tau'},\omega)$ and permeability $\bar{\mu}^{\rho'\sigma'}(\bar{x}^{\tau'},\omega)$ given in the original coordinate system $\mathcal{O}\bar{x}^1\bar{x}^2\bar{x}^3$. Then the corresponding material parameters in the new coordinate system $\mathcal{O}x^1x^2x^3$ are obtained from

$$\varepsilon^{\rho\sigma}(x^{\tau},\omega) = \sqrt{g(x^{\tau})}\; \bar{\Lambda}^{\rho}_{\ \rho'}(x^{\tau})\, \bar{\Lambda}^{\sigma}_{\ \sigma'}(x^{\tau})\; \bar{\varepsilon}^{\rho'\sigma'}(\bar{x}^{\tau'}(x^{\tau}),\omega)\,, \tag{2.86a}$$

$$\mu^{\rho\sigma}(x^{\tau},\omega) = \sqrt{g(x^{\tau})}\; \bar{\Lambda}^{\rho}_{\ \rho'}(x^{\tau})\, \bar{\Lambda}^{\sigma}_{\ \sigma'}(x^{\tau})\; \bar{\mu}^{\rho'\sigma'}(\bar{x}^{\tau'}(x^{\tau}),\omega)\,, \tag{2.86b}$$

where we explicitly noted the usually omitted spatial and frequency dependence for clarification. We see that even if the initial material parameters are isotropic $\bar{\varepsilon}^{\rho'\sigma'} = \bar{\varepsilon}\,\delta^{\rho'}_{\sigma'}$, $\bar{\mu}^{\rho'\sigma'} = \bar{\mu}\,\delta^{\rho'}_{\sigma'}$ the resulting tensors in the new coordinate system

$$\varepsilon^{\rho\sigma} = \sqrt{g}\;\bar{\Lambda}^{\rho}{}_{\rho'}\,\bar{\Lambda}^{\sigma}{}_{\rho'}\,\bar{\varepsilon} = \sqrt{g}\,g^{\rho\sigma}\,\bar{\varepsilon}\,, \tag{2.87a}$$

$$\mu^{\rho\sigma} = \sqrt{g}\;\bar{\Lambda}^{\rho}{}_{\rho'}\,\bar{\Lambda}^{\sigma}{}_{\rho'}\,\bar{\mu} = \sqrt{g}\,g^{\rho\sigma}\,\bar{\mu}\,, \tag{2.87b}$$

will be anisotropic if the mesh is distorted. Here, we used the metric tensor or rather inverse metric tensor which in general curvilinear coordinates can be obtained from the Cartesian metric $\bar{g}_{\rho'\sigma'} = \bar{g}^{\rho'\sigma'} = \delta^{\rho'}_{\sigma'}$ by the transformation rule Eq. (2.76) to be

$$g_{\rho\sigma} = \Lambda^{\rho'}_{\rho}\,\Lambda^{\rho'}_{\sigma}\,, \tag{2.88a}$$

$$g^{\rho\sigma} = \bar{\Lambda}^{\rho}{}_{\rho'}\,\bar{\Lambda}^{\sigma}{}_{\rho'}\,, \tag{2.88b}$$

respectively.

Besides the material properties the focus must be put on the electromagnetic fields. First, the initial fields $\bar{E}_{\rho'}$ and $\bar{H}_{\rho'}$ have to be transformed into the adapted coordinate system according to the transformation rule for covariant vector components

$$E_\rho = \Lambda^{\rho'}_{\rho}\,\bar{E}_{\rho'}\,, \qquad\qquad H_\rho = \Lambda^{\rho'}_{\rho}\,\bar{H}_{\rho'}\,. \tag{2.89}$$

Then, after the problem is solved, the obtained solutions may be transformed back into the original coordinate system with

$$\bar{E}_{\rho'} = \bar{\Lambda}^{\rho}{}_{\rho'}\,E_\rho\,, \qquad\qquad \bar{H}_{\rho'} = \bar{\Lambda}^{\rho}{}_{\rho'}\,H_\rho\,. \tag{2.90}$$

This procedure is to be understood as a transformation for the continuous real space representation of the vector fields from one coordinate system to another coordinate system only. In particular, the transformation steps involved with a discretized representation of the problem and fields suitable for the computational treatment are not included. This might, for example, involve a transformation of an expansion basis which will be discussed later on where appropriate.

The poynting vector, Eq. (2.23), in covariant notation reads

$$S^\rho = \frac{1}{2\sqrt{g}}\,\mathrm{Re}\left(\epsilon^{\rho\sigma\tau}E_\sigma H^*_\tau\right) \tag{2.91}$$

where we used Eq. (2.66) and Eq. (2.80) with positive sign.

With this broad overview of Maxwell's equations in covariant formulation and the transformation rules concerned with arbitrary curvilinear coordinate transformations of the involved tensor quantities, we are finally equipped with all necessary foundations for a rigorous treatment of classical electromagnetic problems.

Chapter 3.
3 Fundamentals of Periodic Systems

In this thesis we are concerned with periodic and artificially periodic nano-photonic structures. The mathematical tools to accurately describe those systems are the topic of this chapter.

In Sec. 3.1 we introduce the mathematical description of periodic systems with the lattice concept in real-space. Section 3.2 covers the corresponding lattice in the dual Fourier space. The Bloch theorem and the Bloch-Floquet expansion as generic solutions for periodic potentials are topic of Sec. 3.3. The transformation procedure between real and Fourier space is discussed in Sec. 3.4, especially the numerical treatment and the occuring obstacles. The last part, Sec. 3.5, recapitulates light diffraction, which occurs at grating-like structures.

3.1. Periodicity and Lattice

The periodicity we are concerned with in the investigated systems is the spatial periodicity of the material. Hence, the periodicity manifests in the optical material parameters permittivity $\underline{\varepsilon}(\mathbf{r})$ and permeability $\underline{\boldsymbol{\mu}}(\mathbf{r})$ in the Euclidean three-dimensional real-space — or direct space. In the following we discuss the permittivity only, but the permeability is assumed to be treated likewise. The periodicity can than be expressed as

$$\underline{\varepsilon}(\mathbf{r}) = \underline{\varepsilon}(\mathbf{r} + \mathbf{R})\,, \tag{3.1}$$

where an arbitrary *lattice vector*

$$\mathbf{R} = l^1\,\mathbf{a}_1 + l^2\,\mathbf{a}_2 + l^3\,\mathbf{a}_3\,, \qquad l^\sigma \in \mathbb{Z}\,, \tag{3.2}$$

which is defined as an integer multiple (l^1, l^2, l^3) of the *lattice basis vectors*[1] $\mathbf{a}_i$, translates the system such that the permittivity remains invariant. The basis is called *primitive* basis. The lattice basis vectors

$$\mathbf{a}_\rho = \sum_\sigma \alpha_\rho^\sigma\,\mathbf{e}_\sigma \tag{3.3}$$

are, in turn, a linear combination of the covariant basis vectors $\mathbf{e}_\rho$ introduced in Sec. 2.5.1. An exemplary lattice is sketched in Fig. 3.1 in the left panel. The (curvilinear) coordinate system is

[1] Note that we often use the term lattice vector instead of lattice basis vector here as well, because every lattice basis vector is also a lattice vector. The context usually clarifies what is actually meant.

usually chosen such that the lattice vectors preferably coincide with the covariant basis vectors, i.e., for a lattice with periodicity along two (not necessarily orthogonal) directions we choose $\mathbf{a}_1 = \alpha_1^1 \mathbf{e}_1$ and $\mathbf{a}_2 = \alpha_2^2 \mathbf{e}_2$. The magnitude of a lattice basis vector $|\mathbf{a}_\rho| := a_\rho$ defines the *lattice constant* in the respective direction.

Please note again, that the lattice constants a_ρ as well as the expansion coefficients α_ρ^σ are scalars and the subscripts ρ are labels which highlight the associated direction. The subscript labels are not to be confused with tensor indices. Also, we advise the reader to recall the notational ambiguity of the superscript σ in the first vector components l^σ and α_ρ^σ alluded to in Sec. 2.5.1. The superscript can mean both, a label or a tensor index, depending whether the quantity is accompanied by a basis vector or not.

The lattice vectors describe the system's periodicity. It is sufficient if the periodic structure is defined within a lattice cell. The translation of a lattice cell with the lattice vectors then covers the whole space without overlap. One distinguishes two types: A lattice *unit cell* is the parallelepiped spanned by the lattice basis vectors [32]. If the lattice basis vectors are primitive, the lattice unit cell is called *primitive cell* and contains only one lattice point. The Wigner-Seitz cell is a special case of a primitive unit cell which inherits the largest symmetry of the lattice [20, 21]. It is defined by the volume around the lattice point that is closer to it than to all other lattice points. If the basis is not primitive, the cell is a *multiple* unit cell and contains more than one lattice point. The second lattice cell type which covers the whole space is called a *conventional cell* [32]. Its basis vectors define a right-handed axial setting, its edges are along symmetry directions of the lattice, and it is the smallest cell compatible with the above condition. Crystals having the same type of conventional cell belong to the same crystal family. Examples for conventional cells are the centered cells, i.e., face-centered-cubic (fcc) or body-centered-cubic (bcc) [20, 21].

In this thesis, we will often work with periodicity and lattices in two dimensions. The third basis vector then points along the direction orthogonal to the plane of periodicity and usually coincides with the z–direction. The predominant lattice type we use is the square lattice, where $\mathbf{a}_1$ and $\mathbf{a}_2$ are perpendicular and of the same length $a_1 = a_2$. What we call unit cell mostly refers to the primitive cell. In order to define the whole structure it suffices to define the permittivity and permeability within this unit cell.

3.2. Reciprocal Space and Reciprocal Lattice

For each lattice in real-space there exists a corresponding lattice in the dual vector space — called the *reciprocal space* or short k–space. The reciprocal lattice is defined by the *reciprocal lattice vectors*

$$\mathbf{G} = m_1 \, \mathbf{a}^1 + m_2 \, \mathbf{a}^2 + m_3 \, \mathbf{a}^3 \,, \qquad m_\sigma \in \mathbb{Z} \,, \tag{3.4}$$

which are an integral linear combination of the dual lattice basis vectors $\mathbf{a}^i$. The dual lattice basis vectors are defined in analogy to Eq. (2.49) or equivalently Eq. (2.51) as

$$\mathbf{a}^\rho = 2\pi \, \frac{\mathbf{a}_\sigma \times \mathbf{a}_\tau}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)} \,, \tag{3.5}$$

with $(\rho, \sigma, \tau) = (1, 2, 3)$ or an even permutation, such that they fulfill the orthogonality relation

$$\mathbf{a}^\rho \, \mathbf{a}_\sigma = 2\pi \delta_\sigma^\rho \,. \tag{3.6}$$

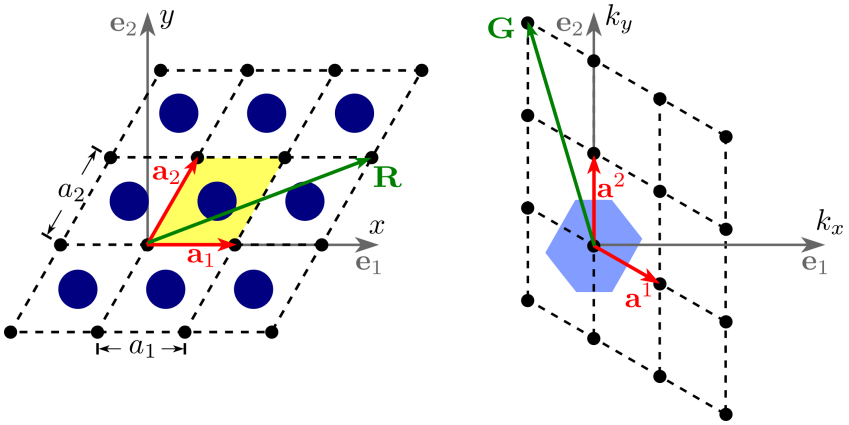


Figure 3.1.: *Example of a direct lattice (left) and a corresponding reciprocal lattice (right). The unit cell in direct space is shaded yellow and contains the structure (dark blue). The first Brillouin-Zone in reciprocal space is shaded in light blue. Two arbitrary lattice and reciprocal lattice vectors are highlighted in green.*

Hence, each reciprocal lattice vector $\mathbf{a}^\rho$ defined by Eq. (3.5) is orthogonal to the two lattice vectors $\mathbf{a}_\sigma$ and $\mathbf{a}_\tau$. The difference to the definition of the dual basis vectors is the phase angle 2π. The larger the lattice constant in direct space, the smaller is the lattice constants in reciprocal space. Consequently, the same is true for the first Brillouin zone (BZ) which is the primitive cell in the reciprocal lattice corresponding to the Wigner-Seitz cell in the direct lattice. An exemplary reciprocal lattice is sketched in Fig. 3.1 in the right panel.

By virtue of the orthogonality condition Eq. (3.6), we can examine a product between a reciprocal lattice vector $\mathbf{G}_m$ and a direct lattice vector $\mathbf{R}_l$ where labels m and l uniquely identify a specific representative of the respective vectors but are usually omitted. The product is given by

$$\begin{aligned}\mathbf{G}_m \cdot \mathbf{R}_l &= (m_1\,\mathbf{a}^1 + m_2\,\mathbf{a}^2 + m_3\,\mathbf{a}^3) \cdot (l^1\,\mathbf{a}_1 + l^2\,\mathbf{a}_2 + l^3\,\mathbf{a}_3) \\ &= 2\pi\,(m_1 l^1 + m_2 l^2 + m_3 l^3) \\ &= 2\pi\,u\,,\end{aligned} \tag{3.7}$$

with $u \in \mathbb{Z}$. The consequence of this integral multiple of a 2π phase is that the exponential term of such product always yields

$$e^{\mathrm{i}\mathbf{G}\mathbf{R}} = e^{\mathrm{i}2\pi u} = 1\,. \tag{3.8}$$

The property described by Eq. (3.8) is the main goal why we introduced the concept of the reciprocal lattice. We deduce from it that a plane wave with k–vector equal to any reciprocal lattice vector $\mathbf{G}$

$$e^{\mathrm{i}\mathbf{G}\,(\mathbf{r}+\mathbf{R})} = e^{\mathrm{i}\mathbf{G}\mathbf{r}} \tag{3.9}$$

reproduces itself after a translation by a vector $\mathbf{R}$. Thus, these plane waves obey the periodicity produced by the structure.

3.3. Bloch-Floquet Theorem

The wave solutions $\Psi(\mathbf{r})$ of a problem with a periodic potential are given by the *Bloch waves*

$$\Psi_{\mathbf{k}}(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r})\, e^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}\,, \tag{3.10}$$

comprising a plane wave envelope with wave vector $\mathbf{k}$, and a lattice periodic amplitude $u_{\mathbf{k}}(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r}+\mathbf{R})$ invariant under the translational symmetry of the potential. This theorem was originally developed by Bloch [33] for the wave functions of electrons as solution to Schrödinger's equation in a periodic potential composed by the regular crystal structure of solids. It is, however, also valid for an electromagnetic wave as solution of Maxwell's equations in a periodic structure, where the periodicity acts as potential for the light field.

Bloch waves have the property that, if shifted by a spatial lattice vector $\mathbf{R}$, they only gain an additional phase factor

$$\Psi_{\mathbf{k}}(\mathbf{r}+\mathbf{R}) = u_{\mathbf{k}}(\mathbf{r}+\mathbf{R})\, e^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}\, e^{\mathrm{i}\mathbf{k}\cdot\mathbf{R}} = \Psi_k(\mathbf{r})\, e^{\mathrm{i}\mathbf{k}\cdot\mathbf{R}}\,, \tag{3.11}$$

which implies that their intensity distribution $\propto |\Psi|^2$ remains fully periodic like the material parameters (if $\mathbf{k} \in \mathbb{R}^n$). A similar translation in reciprocal space,

$$\Psi_{\mathbf{k}+\mathbf{G}_m}(\mathbf{r}) = \Psi_{\mathbf{k}}(\mathbf{r})\,, \tag{3.12}$$

by a reciprocal lattice vector $\mathbf{G}_m$ leaves the wave even totally invariant [34]. Hence, it is sufficient to consider the wave solutions only in the first Brillouin zone, because all solutions with exterior wave vectors can be folded back by Eq. (3.12).

Using Bloch's theorem with the electromagnetic wave equations, Eqs. (2.24), for periodic systems, we find a periodicity in the frequency

$$\omega(\mathbf{k}) = \omega(\mathbf{k}+\mathbf{G}_m) \tag{3.13}$$

with respect to the reciprocal lattice vectors as well. This allows for the back-folding of the dispersion curves into the first BZ without loss of information. Labeling the frequencies with the label of the used reciprocal lattice vector, we notice that for each $\mathbf{k}$–vector within the first BZ there exists an infinite number of solutions with frequencies $\omega_m(\mathbf{k})$.

An important aspect for the purpose of this work is that the periodic amplitude can be expressed by a Fourier series

$$u_{\mathbf{k}}(\mathbf{r}) = \sum_m \tilde{u}_{\mathbf{k},\mathbf{G}_m}\, e^{\mathrm{i}\mathbf{G}_m\mathbf{r}} \tag{3.14}$$

in the reciprocal lattice vectors. Substituting Eq. (3.14) into Eq. (3.10) leads to the Floquet-Fourier expansion[2]

$$\Psi_{\mathbf{k}}(\mathbf{r}) = \sum_m \tilde{u}_{\mathbf{k},\mathbf{G}_m}\, e^{\mathrm{i}(\mathbf{k}+\mathbf{G}_m)\mathbf{r}}\,, \tag{3.15}$$

where the sum over m is infinite and the expansion coefficients $\tilde{u}_{\mathbf{k},\mathbf{G}_m}$ are obtained by a lattice Fourier transformation of the amplitude function. The Floquet-Fourier expansion will become important for the expansion of the electromagnetic fields in the FMM in Chap. 6. Before we get there, the next section introduces Fourier transformations as a basic tool in periodic systems, and explains its necessity and usefulness.

[2] Also known as Bloch-Floquet expansion.

3.4. Fourier Transformations

In the previous sections we have introduced the concepts of direct lattice and reciprocal lattice for periodic systems. Direct and reciprocal space are in general connected with the *Fourier integral transformation* of function $f(\mathbf{r})$ [35]

$$\tilde{f}(\mathbf{k}) = \int_{-\infty}^{\infty} \mathrm{d}\mathbf{r}\, f(\mathbf{r})\, e^{-\mathrm{i}\mathbf{k}\cdot\mathbf{r}} \tag{3.16a}$$

and the inverse transformation

$$f(\mathbf{r}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}\mathbf{k}\, \tilde{f}(\mathbf{k})\, e^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} \tag{3.16b}$$

which makes it easy to switch from one space to the other by transforming the relevant quantities. This transformation is valid for any integrable function which fulfills the Dirichlet conditions[3].

3.4.1. Lattice Fourier Transformation of Periodic Functions

In the special case of periodic functions like the permittivity, Eq. (3.1), or the periodic amplitudes of the associated Bloch waves, Eq. (3.14), the integral Fourier transform reduces to the infinite *Fourier series*, because only k–vectors which comply with the periodicity, the reciprocal lattice vectors $\mathbf{G}_m$, contribute. This means, the permittivity in real-space can be expanded as

$$\underline{\varepsilon}(\mathbf{r}) = \sum_{m=1}^{\infty} \tilde{\underline{\varepsilon}}_m\, e^{\mathrm{i}\,\mathbf{G}_m\cdot\mathbf{r}}\,, \tag{3.17}$$

where the *Fourier coefficients* $\tilde{\underline{\varepsilon}}_m$ are obtained from the *lattice Fourier transform*

$$\tilde{\underline{\varepsilon}}_m = \frac{1}{V_{UC}} \int_{UC} \mathrm{d}\mathbf{r}\, \underline{\varepsilon}(\mathbf{r})\, e^{-\mathrm{i}\,\mathbf{G}_m\cdot\mathbf{r}}\,, \tag{3.18}$$

with the unit cell volume V_{UC}, and $m \equiv (m_1, m_2, m_3)$ labels a specific reciprocal lattice vector as before. For the lattice Fourier transformation of general periodic functions we substitute $\underline{\varepsilon}(\mathbf{r}) \rightarrow f(\mathbf{r})$ and $\tilde{\underline{\varepsilon}}_m \rightarrow \tilde{f}_m$. Equation (3.18) provides the permittivity values at the lattice points of the reciprocal lattice. The integration limits in Eq. (3.18) are chosen to cover the whole unit cell, but are by convention asymmetric to the origin. A symmetric choice around zero would lead to additional phase factors but would not change the result in principle.

[3] The Dirichlet conditions require that the definition interval can be divided into a finite number of intervals in which the function is continuous and monotone. At every discontinuity at the boundary between subsequent intervals the right and left limits must exist [35].

3.4.2. Numerical Treatment

Truncation of the Fourier Series

The expansion of periodic functions into an infinite Fourier series as described in Sec. 3.4.1 cannot be performed on a computer. For a numerical treatment the infinite series has to be truncated to a finite number of M expansion terms. The resulting truncated Fourier series

$$f_M(\mathbf{r}) = \sum_{m=1}^{M} \tilde{f}_m \, e^{\mathrm{i}\,\mathbf{G}_m \cdot \mathbf{r}} \approx f(\mathbf{r}) \tag{3.19}$$

is always[4] an approximation to the original function $f(\mathbf{r})$. The introduced error becomes minimal if the Fourier coefficients obtained by the integral transform Eq. (3.18) are used as expansion coefficients $\tilde{f}_m$ [35]. As long as $f(\mathbf{r})$ is bounded and piecewise continuous, it is also guaranteed that the Fourier series converges to the original function $f_M(\mathbf{r}) \to f(\mathbf{r})$ for $M \to \infty$.

The question that arises is which Fourier coefficients do we include in our series and which are of minor influence and therefore most likely negligible. This question is hard to answer in general because it decisively depends on the shape of the examined structure. Empirical studies show that there are a few rules of thumb that should be respected in order to achieve fast convergence:

1. The coefficients closest to the origin are more important than those farther apart.
2. The truncation scheme should reflect the shape of the structure.

Later on, we will need Fourier series expansions in two periodic dimensions. Therefore, we introduce two simple but general truncation schemes in reciprocal space which are implemented in our framework. The schemes are illustrated in Fig. 3.2.

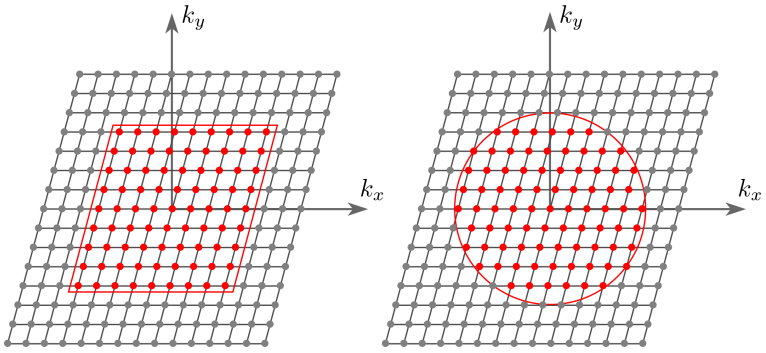


Figure 3.2.: *Parallelogrammic (left) and circular (right) k-space truncation schemes. Both truncations select $M = 81$ k-vectors symmetrically around the origin.*

The *circular truncation* scheme includes contributions from reciprocal lattice points within a circle around the origin $m = (m_1, m_2) = (0, 0)$ of specific radius. The radius is usually determined such that the circle includes $M \leq M_{\text{user}}$ lattice points, where M_{user} is the target value specified by the user.

[4] Except for bandwidth limited functions, of course.

In the *parallelogrammic truncation* scheme, the user defines the maximal orders M_1 and M_2 along both directions defined by the reciprocal lattice basis vectors, and all reciprocal lattice cites m with $-M_1 \leq m_1 \leq +M_1$ and $-M_2 \leq m_2 \leq +M_2$ are included in the series expansion.

For the remainder of this thesis we assume that the Fourier series are properly truncated to allow for numerical treatment and omit the explicit reference to the limits. With this truncation we have made the Fourier series, Eq. (3.19), treatable for the computer. However, the Fourier coefficients of the expansion must still be calculated from a continuous integral equation, Eq. (3.18). The task of the next two paragraphs is to discretize this integral equation and provide an efficient algorithm for the approximation of the Fourier coefficients.

Sampling and Mesh

Numerical calculations can evaluate continuous functions like the kernel $f(\mathbf{r})$ of the integral in Eq. (3.18) only at a finite number of points $\mathbf{r}_i$. This discrete evaluation is called sampling and the evaluation points are called the *sampling points*. The set of sampling points $\{\mathbf{r}_i\colon i = 1, \ldots, N\}$ can be considered as the vertices of a *mesh* (or grid). The meshes commonly used in numerical methods cover a finite, simply connected region of the considered (position) space. Besides the intepretation of a mesh as a collection of sampling points, the vertices are often considered as corners of polyhedra (polygons in 2D) which partition the space into smaller elemental volumes (areas). We would like to stress, that in the context of this work a mesh is interpreted in the former way.

There are two main classifications of meshes: structured and unstructured. An unstructured (or irregular) grid is a tessellation of the considered region by simple shapes, such as triangles or tetrahedra, in an irregular pattern [36,37]. They require a list which specifies the way a given set of vertices make up individual elements (or cells) such that the collection of all individual elements covers the whole space. Furthermore, this list must specify which elements are connected to each other (connectivity). A structured (or regular) grid is a tessellation of the considered region by parallelotopes [37, 38]. Each cell in the grid can be addressed by serial indices from which the connectivity can be simply deduced. This restricts the element choices to quadrilaterals or hexahedra.

The meshes we use in this thesis are two-dimensional regular grids which consist of quadrilateral elements. However, we consider the mesh as regular collection of sampling points at which our functions are evaluated. A special type of such meshes is the equidistant or *Cartesian mesh*. By this we describe a mesh with identical rectangular elements. This means we can describe the coordinates of the sampling points

$$\chi_{kl} = (x_k^1, x_l^2)\,, \qquad x_k^1 = k \cdot \Delta_1\,, \quad x_l^2 = l \cdot \Delta_2\,, \qquad k, l = 0, 1, \ldots, N\,, \tag{3.20}$$

by two indices k and l, and two constant spacings Δ_1 and Δ_2.

The Fast Fourier Transformation

The Fast Fourier Transformation (FFT) is an efficient numerical algorithm to calculate an approximation to the discrete Fourier coefficients $\tilde{f}_m$ on a Cartesian mesh [39].

In order to keep the illustration simple, we present a one-dimensional discrete Fourier transformation

(DFT). If the *equidistant* sampling points are described by N_{fft} coordinates

$$x_k = k \cdot \Delta\,, \qquad k = 0, \ldots, N_{\text{fft}} - 1\,, \tag{3.21a}$$

with sampling interval

$$\Delta = a/N_{\text{fft}}\,, \tag{3.21b}$$

and lattice constant a, then the Fourier coefficients can be approximated as

$$\tilde{f}_m = \frac{1}{a}\int_0^a \mathrm{d}x\, f(x)\, e^{-\mathrm{i}\, m \frac{2\pi}{a}\cdot x} \quad \approx \quad \frac{1}{N_{\text{fft}}} \sum_{k=0}^{N_{\text{fft}}-1} f(x_k)\, e^{-\mathrm{i}\, m \frac{2\pi}{a}\cdot x_k}\,. \tag{3.22}$$

For $N_{\text{fft}} \to \infty$ the sum converges towards the value of the Fourier coefficient.

The FFT algorithm calculates all N_{fft} Fourier coefficients $\tilde{f}_m$ at once using a recursive divide and conquer strategy. For a detailed description we recommend Ref. [39]. Hence, the algorithm is efficient if $N_{\text{fft}} = 2^p$, $p \in \mathbb{N}$. In our code we use the implementation of the open source C subroutine library FFTW (Fastest Fourier Transform in the West) [40, 41], which also incorporates other advanced DFT algorithms, e.g., for numbers of sampling points with small primary factors other than two. The package provides routines for DFTs in one or more dimensions, of arbitrary input size, and of both real and complex data. The FFTW driver routines usually decide with respect to the provided input data about the appropriate algorithmic strategy autonomously. On Intel machines our code uses the signature compatible CPU-optimized FFT routines of the Intel Math Kernel Library (MKL) instead [42].

The biggest advantage of the FFT is that it is rather cheap and that it scales with $\mathcal{O}(NlogN)$ instead of $\mathcal{O}(N^2)$ like an ordinary DFT. The inverse FFT can even be used to quickly calculate the truncated Fourier series of Eq. (3.19). However, there are some fundamental limitations to the accuracy that arise from the discrete sampling which will be discussed in the subsequent paragraph.

Aliasing and Oversampling

Connected to the size of the sampling interval Δ is the so-called *critical frequency* [39]

$$\mathfrak{f}_c = \frac{1}{2\Delta} = \frac{N_{\text{fft}}}{2a}\,. \tag{3.23}$$

Sampling of a sinusoidal wave at that frequency leads to only two sample points per cycle. Thus, waves with higher frequencies than $\mathfrak{f}_c$ are sampled with less than two points per cycle and the sampled function values cannot unambiguously be related to that wave anymore.

The *sampling theorem* [43] states that a bandwidth limited function $f(x)$ with maximal frequency $|\mathfrak{f}_{\max}| \leq \mathfrak{f}_c$ is completely determined by its samples $f(x_k)$. However, if the function is not bandwidth limited, e.g., if the function is discontinuous, then the part of the spectrum outside the frequency interval $(-\mathfrak{f}_c, \mathfrak{f}_c)$ is spuriously moved into that interval and adds up to the coefficients' magnitude there [39]. This effect is called *aliasing*.

Since we truncate the Fourier series anyway, there is a way to decrease the aliasing effect for the retained Fourier coefficients. Looking at the critical frequency, Eq. (3.23), we notice that it is proportional to the number of sampling points N_{fft} via the sampling interval, Eq. (3.21b). Thus, by

increasing the number of sampling points we can increase the critical frequency and reduce the aliasing effect since the spectral components usually decay with increasing distance from the origin. Taking more sampling points N_{fft} than required Fourier coefficients M is called *oversampling*.

The number of sampling points we use by default in calculations is $N_{\mathrm{fft}} = 2^{10} = 1024$ per dimension. This ensures for two-dimensional transformations, where M is usually in the order of 1000 and consequently at the maximum 35 Fourier coefficients (circular truncation) per dimension are needed, that we have a sufficient oversampling factor. If the number of required Fourier coefficients increases and the oversampling drops below a minimum factor of five, we increase N_{fft} successively by factors of two to restore an appropriate oversampling.

We would like to point out that even though the truncated Fourier series approximately reproduces the original functions at the sampling points, it oscillates in between them with the frequency corresponding to the retained Fourier coefficient of highest order.

3.4.3. Gibbs' Phenomenon

The Gibbs' phenomenon describes the fact that a finite set of infinitely differentiable basis functions, like plane waves, can not accurately represent non-differentiable functions [44–46]. Such non-differentiable functions are for example piecewise continuously differentiable periodic permittivity and permeability material functions which have jump discontinuities at material interfaces. This effect is characterized by a ringing — an oscillatory over- and undershoot — at the jump location, whose frequency increases and spatial width decreases with the truncation order, but whose amplitude remains at about nine percent independent of the size of the basis. The Gibbs' phenomenon is illustrated in Fig. 3.3.

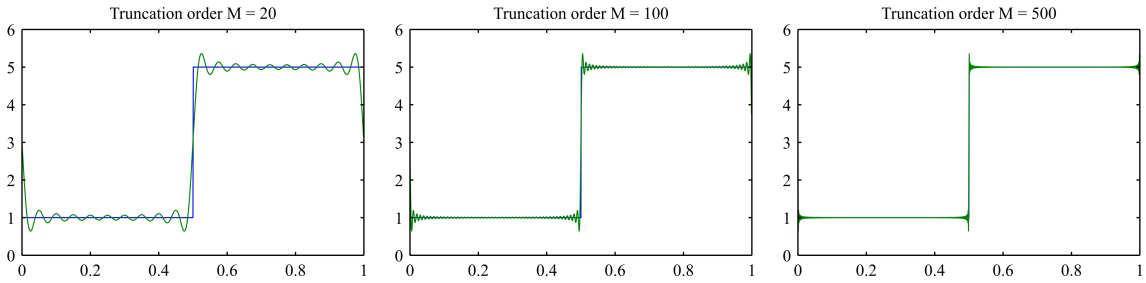


Figure 3.3.: *Illustration of the Gibbs' Phenomenon. The ringing remains for an increasing truncation order. The used parameters are $N_{\mathrm{fft}} = 10000$, $N = 1000$ plot points, and truncation order M as noted above the respective figure.*

3.4.4. Convolution Theorem and Li's Product Factorization Rules

In Sec. 3.4.1 and Sec. 3.4.2 we have discussed how periodic functions can be transformed into the reciprocal space and which discretizations and truncations have to be made in order to describe them numerically.

Especially Maxwell's equations not only consist of functions, but of *products* of periodic functions as, for example, the electric displacement field $\mathbf{D}(\mathbf{r}) = \underline{\varepsilon}(\mathbf{r}) \cdot \mathbf{E}(\mathbf{r})$ which, as a product of permittivity

and electric field, may have jump discontinuities at material interfaces. In reciprocal space products become convolutions [35], i.e., the Fourier transform (denoted with $\mathcal{F}\{\cdot\}$ or a $\tilde{}$ superscript) of the generic function $h(\mathbf{r}) = f(\mathbf{r}) \cdot g(\mathbf{r})$ is

$$\mathcal{F}\{h\} = \mathcal{F}\{\, f \cdot g \,\} = \mathcal{F}\{f\} * \mathcal{F}\{g\}\,, \tag{3.24a}$$

where the convolution of the continuous transformed functions is given by

$$\tilde{h}(\mathbf{k}) = \left[\, \tilde{f} * \tilde{g} \,\right](\mathbf{k}) = \int_{-\infty}^{+\infty} \mathrm{d}\mathbf{k}'\, \tilde{f}(\mathbf{k} - \mathbf{k}')\, \tilde{g}(\mathbf{k}')\,. \tag{3.24b}$$

In the discretized k–space of the reciprocal lattice the convolution integral transforms into

$$\tilde{h}_m^{(N)} = \sum_{n=1}^{N} \tilde{f}_{m-n}\, \tilde{g}_n\,, \tag{3.25}$$

which is a sum over the product of those Fourier coefficients of f and g labeled with multi-indices $n = (n_1, n_2, n_3)$ (equivalent to m) and $m - n = (m_1 - n_1, m_2 - n_2, m_3 - n_3)$, such that the sum of the corresponding reciprocal lattice vectors on the right hand side of the equation is identical to the requested Fourier coefficient $m = (m_1, m_2, m_3)$ on the left. We notice again, that the sum is usually infinite $(N \to \infty)$ but must be truncated to a finite number of terms N in order to be calculated by a computer. A common choice we adopt for the remainder of this work is to set $N = M$ and pick the same reciprocal lattice vectors symmetrically around the origin and in the same order for multi-indices m and n. Which reciprocal lattice vectors the set of M coefficients contains depends on the truncation scheme (cf. Sec. 3.4.2). However, note that if the maximum contained order in the set is given by $m_{\rho,\max} = \max(|m_\rho|) = n_{\rho,\max}$, the difference $\max(|m_\rho - n_\rho|) = 2m_{\rho,\max}$ requires coefficients up to twice the maximal order. This is a second goood reason for oversampling (cf. Sec. 3.4.2). Hence, the product factorization Eq. (3.25) — which is called *Laurent's rule* — can be written in matrix–vector notation

$$\tilde{\mathbf{h}} = [\![f]\!]\, \tilde{\mathbf{g}}\,, \tag{3.26}$$

where $\tilde{\mathbf{h}}$ and $\tilde{\mathbf{g}}$ are vectors containing the respective Fourier coefficients, and the coefficients $\tilde{f}_{m-n}$ are the (m, n)-th entry of matrix $[\![f]\!]$. For a one-dimensional transformation, this matrix has Toeplitz structure (cf. e.g. Ref. [39]). Unfortunately, for a two-dimensional transformation, which is what we will use later on, the matrix only has block Toeplitz structure with usually small blocks. The exact shape depends on the sequence of retained reciprocal lattice vectors corresponding to the multi-indices m and n.

One-Dimensional Factorizations

The truncated Fourier series, Eq. (3.19), with the finite Fourier factorization of the product $f \cdot g$ introduced in Eq. (3.25)

$$h^{(M)}(x) = \sum_{m=1}^{M} \tilde{h}_m^{(M)}\, e^{\mathrm{i}\, G_{1,m} \cdot x} \tag{3.27}$$

converges towards the original function

$$h^{(M)}(x) \to h(x)\,, \qquad \text{for } M \to \infty\,, \tag{3.28}$$

only if certain requirements are met. Li [47,48] established rules for different types of functions and proved his theorems at least for one-dimensional transformations[5].

Li's rules state: A product of two piecewise-smooth, bounded, periodic functions that have ...

- **Type 1:** ... *no concurrent* jump discontinuities can be Fourier factorized by Laurent's rule, Eq. (3.25),
- **Type 2:** ... only *pairwise complementary* jump discontinuities can be Fourier factorized by the *inverse rule*, which replaces the Toeplitz matrix of function f by the inverse Toeplitz matrix of the function $1/f$,

$$\tilde{\mathbf{h}} = \left[\!\left[1/f\right]\!\right]^{-1} \tilde{\mathbf{g}}\,, \tag{3.29}$$

- **Type 3:** ... *concurrent but not complementary* jump discontinuities can be Fourier factorized neither by Laurent's rule nor by the inverse rule,

such that Eq. (3.28) is valid.

The products occurring in Maxwell's equations are either of type 1 or of type 2. Most products are factorized with Laurent's rule. Only for the electric displacement $\mathbf{D}$ and the magnetic induction $\mathbf{B}$ the correct rule depends on the continuity conditions Eq. (2.16) which require a factorization using the inverse rule when the Fourier transformation is carried out along a direction of continuous fields. A complication of the procedure occurs when the Fourier transformation is performed in two dimensions. This is the topic of the next paragraph.

Two-Dimensional Factorizations

Let us now consider the more practical case of a lattice Fourier transformation of the linear constitutive relations, Eq. (2.8), appearing in covariant Maxwell's equations as

$$D^{\rho}(x^1, x^2) \;=\; \varepsilon^{\rho\sigma}(x^1, x^2)\; E_{\sigma}(x^1, x^2)\,, \tag{3.30a}$$

$$B^{\rho}(x^1, x^2) \;=\; \mu^{\rho\sigma}(x^1, x^2)\; H_{\sigma}(x^1, x^2)\,, \tag{3.30b}$$

along the two directions of periodicity x^1 and x^2. The dielectric displacement and magnetic induction Fourier factorize in exactly the same way, which is why we restrict our illustration of the procedure to the former. In the following, the spatial dependence is assumed but mostly suppressed for brevity.

We will perform Fourier transformations in dimensions x^1 and x^2 successively one after the other. Li's rules describe which transformation rules should be chosen for the transformation of products of functions, namely Laurent's rule, Eq. (3.25) or Eq. (3.26), for products of at most one discontinuous function at a certain spatial point, and the inverse rule, Eq. (3.29), for functions with concurrent complementary jump discontinuities. The permittivity exhibits jump discontinuities at material interfaces, whereas some of the field components show a concurrent complementary jump and others are continuous at the interfaces.

[5] Li actually only gave a very rough sketch of the proof (cf. Ref. [48]). Branimir Anic from the Mathematical Department of KIT carried out the whole proof in one dimension with our support and filled Li's gaps [49]. It is several pages long. He currently works on the proof for two-dimensional transformations, which is even more cumbersome.

The full derivation of the factorization is carried out in App. A.1 and reworks the findings of Li [50, 51]. Here, we sketch the general idea only by means of the first component

$$D^1 = \varepsilon^{11} E_1 + \varepsilon^{12} E_2 + \varepsilon^{13} E_3 \,. \tag{3.31}$$

Let us start with a Fourier transformation in x^1 direction. Thus, in Eq. (3.31) we know for sure that according to the continuity conditions D^1, E_2, and E_3 must be continuous across the whole unit cell.[6] The first step is to reformulate the equation in such a way that we only get products of functions which are of Laurent or inverse type. The appropriate reformulation reads

$$D^1 = \varepsilon^{11} \left[E_1 + \left(\frac{\varepsilon^{12}}{\varepsilon^{11}} \right) E_2 + \left(\frac{\varepsilon^{13}}{\varepsilon^{11}} \right) E_3 \right] \,. \tag{3.32}$$

Since D^1 is continuous, but ε^{11} is definitely discontinuous at material interfaces, the term in the square brackets must also be discontinuous and its product with ε^{11} must be of inverse type. The products within the square brackets are of Laurent type because E_2 and E_3 are continuous. This implies

$$\begin{aligned} \tilde{\mathbf{D}}^1 &= \left[\!\left[\frac{1}{\varepsilon^{11}} \right]\!\right]^{-1} \tilde{\mathbf{E}}_1 + \left[\!\left[\frac{1}{\varepsilon^{11}} \right]\!\right]^{-1} \left[\!\left[\frac{\varepsilon^{12}}{\varepsilon^{11}} \right]\!\right] \tilde{\mathbf{E}}_2 + \left[\!\left[\frac{1}{\varepsilon^{11}} \right]\!\right]^{-1} \left[\!\left[\frac{\varepsilon^{13}}{\varepsilon^{11}} \right]\!\right] \tilde{\mathbf{E}}_3 \\ &= \quad \underline{\mathbf{Q}}^{11} \, \tilde{\mathbf{E}}_1 \quad + \quad \underline{\mathbf{Q}}^{12} \, \tilde{\mathbf{E}}_2 \quad + \quad \underline{\mathbf{Q}}^{13} \, \tilde{\mathbf{E}}_3 \,, \end{aligned} \tag{3.33}$$

where all field components and matrices[7] still depend on x^2.

Next, we perform the Fourier transformation along x^2-direction. Then, the fields D^2, E_1, and E_3 must be continuous across the whole unit cell. Consequently, we do not know how $\tilde{\mathbf{D}}^1$ and $\tilde{\mathbf{E}}_2$ behave, and we must eliminate one of them. To this end, we substitute for $\tilde{\mathbf{E}}_2$ the reordered, x^1–transformed component $\tilde{\mathbf{D}}^2$ (cf. Eq. (A.9))

$$\tilde{\mathbf{E}}_2 = \left(\underline{\mathbf{Q}}^{22} \right)^{-1} \tilde{\mathbf{D}}^2 \; - \; \left(\left(\underline{\mathbf{Q}}^{22} \right)^{-1} \underline{\mathbf{Q}}^{21} \right) \tilde{\mathbf{E}}_1 \; - \; \left(\left(\underline{\mathbf{Q}}^{22} \right)^{-1} \underline{\mathbf{Q}}^{23} \right) \tilde{\mathbf{E}}_3 \,, \tag{3.34}$$

and get after some algebraic transformations

$$\begin{aligned} \tilde{\mathbf{D}}^1 \; &= \; \left(\underline{\mathbf{Q}}^{11} - \underline{\mathbf{Q}}^{12} \left(\underline{\mathbf{Q}}^{22} \right)^{-1} \underline{\mathbf{Q}}^{21} \right) \tilde{\mathbf{E}}_1 \; + \; \left(\underline{\mathbf{Q}}^{12} \left(\underline{\mathbf{Q}}^{22} \right)^{-1} \right) \tilde{\mathbf{D}}^2 \\ &\quad + \left(\underline{\mathbf{Q}}^{13} - \underline{\mathbf{Q}}^{12} \left(\underline{\mathbf{Q}}^{22} \right)^{-1} \underline{\mathbf{Q}}^{23} \right) \tilde{\mathbf{E}}_3 \\ &=: \; \check{\underline{\mathbf{Q}}}^{11} \, \tilde{\mathbf{E}}_1 \; + \; \check{\underline{\mathbf{Q}}}^{12} \, \tilde{\mathbf{D}}^2 \; + \; \check{\underline{\mathbf{Q}}}^{13} \, \tilde{\mathbf{E}}_3 \,, \end{aligned} \tag{3.35}$$

where the field components on the right hand side are all continuous. Hence, all terms on the right hand side are Fourier factorized according to Laurent's rule. The final step to the Fourier transformed first component of the linear electric constitutive relation is to resubstitute component $\tilde{\tilde{\mathbf{D}}}^2$ by

[6] To be more precise, the continuity conditions cannot be fulfilled at (sharp) corners of material interfaces. There, the fields diverge [52, 53]. But these points are usually very few. Furthermore, the singularities can not be well represented in a truncated Fourier series. The field representation at those singular points are a principle problem for many numerical methods. Therefore, we ignore them in our considerations (cf. Ref [51]).

[7] More precisely, even every single matrix entry is a function of x^2

Eq. (A.11) which results in

$$\begin{aligned}
\tilde{\tilde{\mathbf{D}}}^1 &= \left[\!\left[\underline{\check{\mathbf{Q}}}^{11}\right]\!\right] \tilde{\tilde{\mathbf{E}}}_1 + \left[\!\left[\underline{\check{\mathbf{Q}}}^{12}\right]\!\right] \tilde{\tilde{\mathbf{D}}}^2 + \left[\!\left[\underline{\check{\mathbf{Q}}}^{13}\right]\!\right] \tilde{\tilde{\mathbf{E}}}_3 \\
&=: \underline{\bar{\mathbf{Q}}}^{11} \tilde{\tilde{\mathbf{E}}}_1 + \underline{\bar{\mathbf{Q}}}^{12} \tilde{\tilde{\mathbf{D}}}^2 + \underline{\bar{\mathbf{Q}}}^{13} \tilde{\tilde{\mathbf{E}}}_3 \\
&\overset{\text{Eq. (A.11)}}{=} \left(\underline{\bar{\mathbf{Q}}}^{11} + \underline{\bar{\mathbf{Q}}}^{12} \left(\underline{\bar{\mathbf{Q}}}^{22}\right)^{-1} \underline{\bar{\mathbf{Q}}}^{21}\right) \tilde{\tilde{\mathbf{E}}}_1 + \left(\underline{\bar{\mathbf{Q}}}^{12} \left(\underline{\bar{\mathbf{Q}}}^{22}\right)^{-1}\right) \tilde{\tilde{\mathbf{E}}}_2 \\
&\quad + \left(\underline{\bar{\mathbf{Q}}}^{13} + \underline{\bar{\mathbf{Q}}}^{12} \left(\underline{\bar{\mathbf{Q}}}^{22}\right)^{-1} \underline{\bar{\mathbf{Q}}}^{23}\right) \tilde{\tilde{\mathbf{E}}}_3 \\
&=: \underline{\tilde{\boldsymbol{\varepsilon}}}^{11} \tilde{\tilde{\mathbf{E}}}_1 + \underline{\tilde{\boldsymbol{\varepsilon}}}^{12} \tilde{\tilde{\mathbf{E}}}_2 + \underline{\tilde{\boldsymbol{\varepsilon}}}^{13} \tilde{\tilde{\mathbf{E}}}_3 \,,
\end{aligned} \tag{3.36}$$

The remaining components are obtained in a similar fashion. They can be found in App. A.1.

There is an important technical remark to make for the Fourier transformation along direction x^2 comprised in matrices $\underline{\bar{\mathbf{Q}}}^{\rho\sigma}$ in Eq. (3.36). Since each of the terms $\underline{\check{\mathbf{Q}}}^{\rho\sigma}$ in Eq. (3.35) is still dependent on x^2, we actually have not a single matrix $\underline{\check{\mathbf{Q}}}^{\rho\sigma}$, but one matrix $\underline{\check{\mathbf{Q}}}^{\rho\sigma}(x_k^2)$ for each sampling point x_k^2, $k = 1, \ldots, N_{\text{fft}}$.

The final quantitiy in Fourier space $\tilde{\tilde{\mathbf{D}}}^1 = (\ldots, \tilde{\tilde{D}}^1_m, \ldots)$ is a vector of Fourier coefficients, where $m = (m_1, m_2)$ is the multi-index labeling one specific order in two-dimensional k-space. There are M orders in total which we take into account ($m = 1, \ldots, M$). Considering Eq. (3.25), we can write

$$\tilde{\tilde{D}}^1_m = \sum_{n=1}^{M} \left(\underline{\bar{\mathbf{Q}}}^{11}\right)_{mn} \tilde{\tilde{E}}^1_n + \ldots \,, \tag{3.37}$$

where $n = (n_1, n_2)$, $n = 1, \ldots, M$ is a second multi-index running over the same k-orders as m.

In order to obtain matrix $\underline{\bar{\mathbf{Q}}}^{\rho\sigma}$, we stack the matrices $\underline{\check{\mathbf{Q}}}(x_k^2)$ into a three dimensional array A where each slice, e.g., the fourth slice $A(m, n, k = 4) = \underline{\check{\mathbf{Q}}}(x_4^2)$, represents one of these Toeplitz matrices (here for $x^2 = x_4^2$). The Fourier transformation in x^2 is then performed for each array entry stack independently, i.e., the transformation of the first entry's stack $A(m = 1, n = 1, k)$ is carried out over the third array index k, and so on. We store the Fourier transformed array $\tilde{A}$ in the same way. The usual FFT output relates $k = 1$ to the zeroth and $k = N_{\text{fft}}$ to the (-1)-st Fourier coefficient (FFT ordering).

We recall that the (m, n)-th entry of matrices $(\underline{\mathbf{Q}}^{\rho\sigma})^{-1}$ contains the $(m_1 - n_1)$-th Fourier coefficient of the respective function — this is the definition of a Toeplitz matrix after all. Similarly, the (m, n)-th entry of matrices $\underline{\bar{\mathbf{Q}}}^{\rho\sigma}$ contains the $(m_2 - n_2)$-th Fourier coefficient. This corresponts to array entry $\tilde{A}(m, n, \tilde{k})$, where $\tilde{k} = \text{mod}(m_2 - n_2 + N_{\text{fft}}, N_{\text{fft}})$ is the index of the $(m_2 - n_2)$-th Fourier coefficient and mod denotes the integer remainder (modulo operator).[8]

[8] This procedure is indeed not easy. Recent investigations [54] indicate that it might not be necessary to distinguish between Laurent and and Inverse rule in the two directions as demanded by Li. We observed that the Inverse rule converges with a similar rate for Type 1 problems as the Laurent rule. This implies the application of the Inverse rule in both dimensions without loss of accuracy. Such a procedure would be much easier and faster than the described scheme. This hypothesis remains yet to be tested in future works.

Li's Transformation Operators

The procedure presented in the last paragraph is complicated and confusing. However, the whole transformation process can be made more clear when we introduce a few operators that systematize and facilitate the rearrangement of the tensor components to form proper type 1 or type 2 products and the Fourier transformation. These operators have been introduced by Li [51] and they separate the procedures in directions 1 and 2. The Fourier transformed permittivity tensor $\underline{\hat{\varepsilon}}$ introduced above is obtained from the real-space permittivity tensor $\underline{\varepsilon}(x^1, x^2)$ by successive application of operators

$$\hat{\mathrm{L}}_\tau = \hat{\mathrm{l}}_\tau^+ \, \hat{\mathrm{F}}_\tau \, \hat{\mathrm{l}}_\tau^- \,, \tag{3.38}$$

for directions $\tau = 1, 2$. The operators $\hat{\mathrm{l}}_\tau^\pm$ rearrange the permittivity tensor elements back and forth into the necessary form for a Fourier factorization like we have for example done manually in Eq. (3.32). For an arbitrary 3×3 tensor[9] $\underline{\mathbf{A}}$ the operators are defined by $\underline{\mathbf{B}} = \hat{\mathrm{l}}_\tau^\pm(\underline{\mathbf{A}})$, with

$$B^{\rho\sigma} = \begin{cases} (A^{\tau\tau})^{-1} \,, & \rho = \tau, \sigma = \tau, \\ (A^{\tau\tau})^{-1} A^{\tau\sigma} \,, & \rho = \tau, \sigma \neq \tau, \\ A^{\rho\tau} (A^{\tau\tau})^{-1} \,, & \rho \neq \tau, \sigma = \tau, \\ A^{\rho\sigma} \pm A^{\rho\tau} (A^{\tau\tau})^{-1} A^{\tau\sigma} \,, & \rho \neq \tau, \sigma \neq \tau. \end{cases} \tag{3.39}$$

The operator $\hat{\mathrm{F}}_\tau$ performs the Fourier transformation with respect to coordinate x^τ and generates the corresponding Toeplitz matrix as described above. With the help of these operators the entire $\underline{\mathbf{Q}}$ tensor, whose elements where introduced in Eq. (3.33) can be written as

$$\underline{\mathbf{Q}}(x^2) = \hat{\mathrm{L}}_1(\underline{\varepsilon}) = \hat{\mathrm{l}}_1^+ \, \hat{\mathrm{F}}_1 \, \hat{\mathrm{l}}_1^-(\underline{\varepsilon}) \,, \tag{3.40}$$

and the permittivity tensor in reciprocal space defined in Eq. (3.36) is given by

$$\underline{\tilde{\varepsilon}} = \hat{\mathrm{L}}_2 \hat{\mathrm{L}}_1(\underline{\varepsilon}) = \hat{\mathrm{l}}_2^+ \, \hat{\mathrm{F}}_2 \, \hat{\mathrm{l}}_2^- \, \hat{\mathrm{l}}_1^+ \, \hat{\mathrm{F}}_1 \, \hat{\mathrm{l}}_1^-(\underline{\varepsilon}) \,. \tag{3.41}$$

The order of the transformation is not unique. We can equally well first transform with respect to x^2 and thereafter along coordinate x^1. The resulting reciprocal permittivity tensor

$$\underline{\tilde{\varepsilon}}' = \hat{\mathrm{L}}_1 \hat{\mathrm{L}}_2(\underline{\varepsilon}) \quad \neq \quad \underline{\tilde{\varepsilon}} \,, \tag{3.42}$$

cannot be derived from the tensor in Eq. (3.41) for a finite truncate Fourier transform [51]. However, the expectation is that for $M \to \infty$ both representations of the permittivity in reciprocal space converge to the same limit. Numerical test show that symmetries in the permittivity distribution are not conserved in either of the k–space representations, but the deviations diminish when the truncation order is increased. To restore the symmetry in reciprocal space Li suggests to use the average of both representations

$$\frac{\underline{\tilde{\varepsilon}} + \underline{\tilde{\varepsilon}}'}{2} \tag{3.43}$$

[9] Note that the tensor elements can be matrices themselves.

instead of using only one of them or picking some components from the one and some from the other. This symmetrization has the disadvantage that the intrinsic perfect energy balance of the FMM (cf. Chap. 6) is lost. But in all practical applications the number of retained Fourier coefficients is large enough so that the deviation from full energy conservation is negligible.

Finally, we have gathered all necessary basic concepts to represent periodic structures in reciprocal space by means of lattice Fourier transformations. For a numerical treatment the infinite series have been truncated and the continuous integrals have been discretized. Li's rules provide the guidlines to dramatically improve the convergence of Fourier factorized products of functions with concurrent, pairwise complementary jump discontinuities, and his transformation operators formalize and facilitate the procedure of applying those rules to the constitutive relations in the correct way.

3.5. Diffraction

The behavior of plane waves at a planar interface between two homogeneous, isotropic half-spaces characterized by their respective refractive indices n_1 and n_2 was discussed in Sec. 2.4.3. Here, we would like to discuss the implications of an additional periodic layer **3** in between the two regions. The situation we are interested in is that of a simple diffraction grating which is schematically depicted in Fig. 3.4 for a periodicity along the x^1–direction, but the derivations we present are for the more general case of two-dimensional periodicity. One dimensional gratings are often called

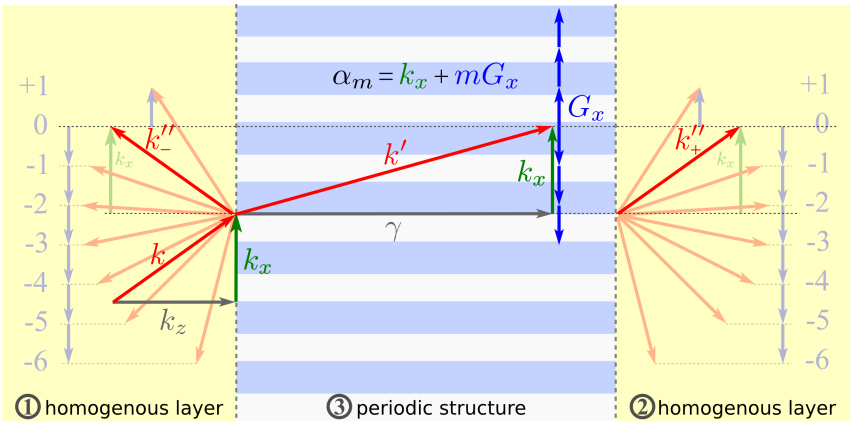


Figure 3.4.: *Sketch of diffraction at a one-dimensional periodic layer and corresponding Bragg orders (x-z-plane) for a wave incident from the left. The wave vector component parallel to the material interface k_x is conserved across the interface. Within the periodic structure reciprocal lattice vectors G_x are added. These are, in turn, conserved across the interface in the transmitted and reflected waves and lead to diffraction into Bragg orders.*

lamellar gratings and two dimensional gratings are called *crossed* gratings. The periodic layer is characterized by its refractive index $n_\mathbf{3} = n_\mathbf{3}(x^1, x^2)$ which defines a lattice with constants a_1 and a_2.

The structure is illuminated with a plane wave with frequency ω and wave vector

$$\mathbf{k} = \alpha_0 \mathbf{a}^1 + \beta_0 \mathbf{a}^2 + \gamma_0 \mathbf{a}^3 \tag{3.44}$$

from region $\mathbf{1}$. We know from Eq. (2.35) that the wave vector components $\mathbf{k}_\parallel$ parallel to the planar interfaces are conserved. However, inside the periodic region, according to Eq. (3.12), there is an infinite number of Bloch waves with the same frequency ω whose in-plane wave vectors differ by a reciprocal lattice vector $\mathbf{G}_m$ which can all be excited. They all exhibit a different $\mathbf{k}_3$ component though. The electromagnetic field inside the structure, thus, consists of Bloch waves, which are a superposition of plane waves with wave vectors

$$\mathbf{k}_\parallel + \mathbf{G}_m \,, \qquad \mathbf{G}_m = m_1 \, \mathbf{a}^1 + m_2 \, \mathbf{a}^2 = m_1 \, \frac{2\pi}{a_1} \mathbf{e}_1 + m_2 \, \frac{2\pi}{a_2} \mathbf{e}_2 \tag{3.45}$$

represented by the Floquet-Fourier series of Eq. (3.15). The excitation strength of the Bloch modes and thereby the single plane waves is determined from the field continuity conditions Eq. (2.16) and Eq. (2.19); the exact distribution is not of interest here. We are rather focusing on the fact that because of the related continuity of the $\mathbf{k}_\parallel$–vector this $\mathbf{G}_m$ modulus is transfered into the reflected and transmitted fields in the homogenous layers.

The reflected and transmitted waves in regions $\mathbf{1}$ and $\mathbf{2}$ do no longer consist of a single plane wave each, as presented for the interface between homogeneous layers in Sec. 2.4.3. Instead, the fields are given by the Rayleigh expansion [55], e.g., the electric field by

$$\mathbf{E_1}(\mathbf{r}) = \mathbf{E}_{\text{in}} \, e^{\mathrm{i}\mathbf{k}\mathbf{r}} + \sum_m \mathbf{E}_{\text{refl},m} \, e^{\mathrm{i}(\alpha_{m_1} x^1 + \beta_{m_2} x^2 - \gamma_m x^3)} \tag{3.46a}$$

in region $\mathbf{1}$, and

$$\mathbf{E_2}(\mathbf{r}) = \sum_m \mathbf{E}_{\text{trans},m} \, e^{\mathrm{i}(\alpha_{m_1} x^1 + \beta_{m_2} x^2 + \gamma_m x^3)} \tag{3.46b}$$

in region $\mathbf{2}$, where $m = \{m_1, m_2\}$ is a multi-index as before which defines the *diffraction order*. The first, second, and third contravariant components of the wave vector $\mathbf{k}_m = \mathbf{k} + \mathbf{G}_m$ corresponding to the m-th diffraction order are given by

$$\alpha_{m_1} = \alpha_0 + m_1 \, \frac{2\pi}{a_1} \,, \tag{3.47a}$$

$$\beta_{m_2} = \beta_0 + m_2 \, \frac{2\pi}{a_2} \,, \tag{3.47b}$$

$$\gamma_m = \gamma_{m_1,m_2} \,, \tag{3.47c}$$

respectively. The latter must be determined from the plane wave dispersion relation, Eq. (2.30),

$$\mathbf{k}_m \cdot \mathbf{k}_m = g^{\rho\sigma} \, k_\rho k_\sigma = n_{\mathbf{s}}^2 \, \omega^2 \,, \qquad \mathbf{s} = \mathbf{1}, \mathbf{2} \,, \tag{3.48}$$

for regions $\mathbf{1}$ and $\mathbf{2}$ separately, where k_ρ is the ρ-th component of $\mathbf{k}_m$. Hence, Eq. (3.48) must be solved for $k_3 = \gamma_m$ for every diffraction order m.

In principle, the sum over m in the Rayleigh expansion, Eq. (3.46), is infinite. However, there is only a finite number of *propagating* diffraction orders which have a real propagation constant γ_m.

The evanescent waves with imaginary propagation constant only contribute to the near fields in the two half-spaces. This is similar to the total internal reflection for an interface between homogeneous layers presented in Sec. 2.4.3. The number of propagating diffraction orders — or Bragg orders, as they are often called — depends on the frequency, the angle of incidence, the lattice constants and the refractive index of the respective homogeneous halfspace. For wavelengths $\lambda = 2\pi/\omega > na_i$, $i = 1, 2$, there are no diffracted plane waves for normal incidence on the grating. Only the reflected and transmitted waves exist (zeroth order). The smaller the ratio between wavelength and lattice constant, the more Bragg orders can be observed.

When taking a spectrum of a grating with given lattice constant, at wavelengths where an additional propagating diffraction order is allowed, a resonance feature in transmittance or reflectance can be noticed. This feature is called Wood's anomaly[10] [56] and can be explained by the redistribution of energy carried by the waves among the increased number of diffraction orders.

[10] A different typical name is Rayleigh anomaly.

Chapter 4.

4 Fundamentals of Optical Waveguides

Optical waveguides are elongated dielectric structures with constant cross section used to direct electromagnetic waves in the visible and near infrared regime of the spectrum [57]. Their typical extents range from a few to hundreds of micrometers in diameter and their lengths are usually much larger and fit to the requirements of the respective field of application. In practice, they are mostly used for optical telecommunications in form of fibers, but in future they will become more and more interesting in the context of integrated optical circuits in form of ridge or strip waveguides.

Physically, their guiding principle is based on total internal reflection of electromagnetic waves at an interface between a material with a high and one with a low permittivity ε (or refractive index n) like introduced in Sec. 2.4.3. Hence, a typical circular fiber consists of a central *core* of high index transparent material such as glass, and a surrounding *cladding* layer with smaller refractive index, e.g., another sort of glass. The cladding is, in turn, surrounded by a mechanically protective *jacket* layer, as schematically depicted in Fig. 4.1. Ridge waveguides are composed similarly. The core

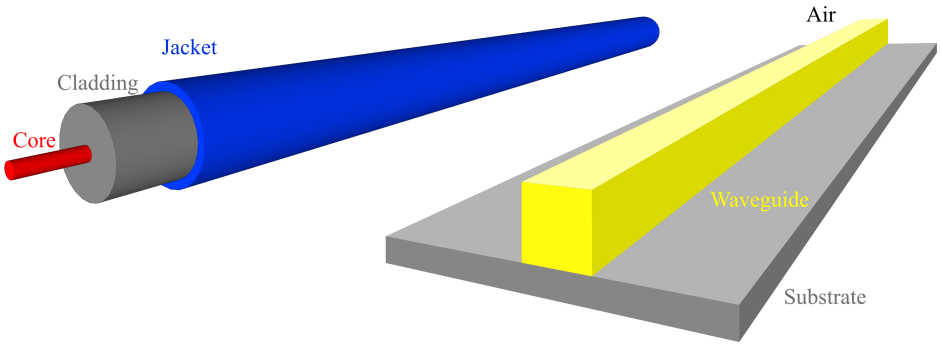


Figure 4.1.: *Conventional step-index fiber waveguide (left), and ridge waveguide (right). Typical fiber diameters: core* $5 - 100\,\mu\mathrm{m}$*, cladding* $125\,\mu\mathrm{m}$*, jacket* $250\,\mu\mathrm{m}$*. Typical ridge dimensions:* $0.2 - 5\,\mu\mathrm{m}$.

is mostly made of silicon or another suitable material available for silicon-based high integration schemes in industrial semiconductor manufacturing. The cladding is often air and/or a low index substrate, e.g., silica SiO_2. For a guiding effect it is sufficient that the index difference is in the order of one percent. The core's cross section is usually rectangular, but for some applications it is interesting to use different shapes, like trapezoids for instance (cf. Fig. 4.2).

In analogy to the quantum mechanical potential well, we can consider the permittivity as "potential" for the electromagnetic waves but with the opposite sign. Instead of the permittivity, we would rather like to use the related refractive index for the discussion of this analogy, because it is the commonly used quantity in context of waveguides. The difference between core and cladding refractive index then corresponds to the depth of the well. Similarly to bound electronic solutions of Schrödinger's equation with discrete quantized energy levels in the potential well, there exists a finite set of bound electromagnetic wave solutions of Maxwell's equations in the core. However, the confinement in the core is only in the transverse direction. Every bound electromagnetic solution travels along the waveguide with an axially wavevector component, which is often referred to as its *propagation constant* β.[1] We introduce the *effective (refractive) index*

$$\mathrm{n_{eff}} \equiv \frac{\beta}{\omega} = c_0 \frac{\beta_{\mathrm{SI}}}{\omega_{\mathrm{SI}}} \tag{4.1}$$

as the analogon to the discrete energy levels in the waveguide system. These effective indices are not regularly spaced. The guided eigenmode of the waveguide structure with the highest effective index is called the *fundamental mode*. The discrete bound states have an oscillatory spatial dependence within the core and decay exponentially in the cladding. Therefore, the majority of the carried power is confined in the waveguide core. Since the cladding is usually sufficiently thick, the influence of the jacket can be neglected in the considerations. Below the *guiding cutoff* — the refractive index of the cladding — there is a continuum of radiative solutions.

A special case in the zoo of structures guiding light by the principle of total internal reflection are slotted optical waveguides. These structures, discovered by Almeida *et al.* in 2004 [58], consist of two ridge waveguides built side by side with a small sub-wavelength wide gap in between, as sketched in Fig. 4.2. According to the continuity conditions, Eq. (2.16), the normal field components

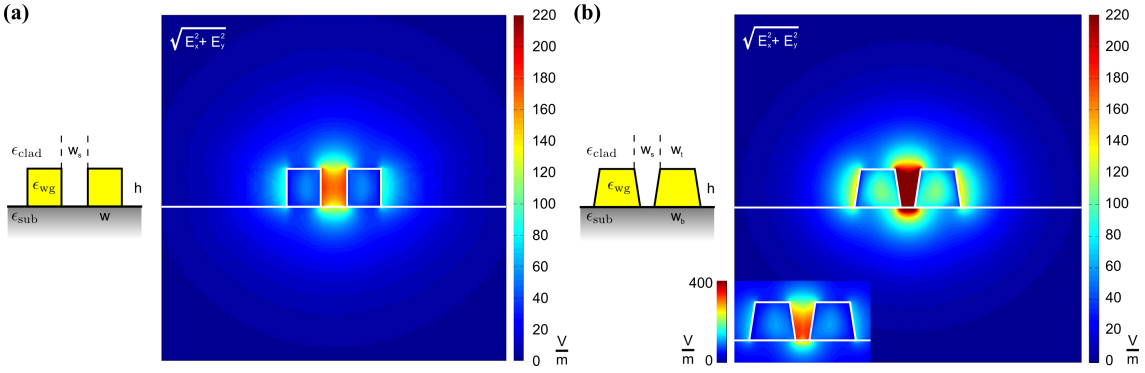


Figure 4.2.: *Schematic and fundamental mode of sub-wavelength slotted waveguides as simulated with the FMM eigenmode solver. (a) Rectangular slotted-waveguide, and (b) trapezoidal slotted-waveguide. The slanted walls of the latter lead to enhanced field intensities and smaller mode volume. Picture and parameters see [59].*

of $\mathbf{D}$ and $\mathbf{B}$ must be continuous across the side wall interface of the ridge waveguides. Hence, when the permittivity of the core is much larger than the permittivity in the gap, the normal field components in the gap must increase antiproportionally in order to fulfill the continuity conditions.

[1] We will later use γ for the propagation constant, as is the custom in FMM literature.

The exponential decay in the gap then occurs on much larger length scales than the width of the gap. Consequently, the independent guided modes of separated ridge waveguides hybridize to bound slot-waveguide modes with maximum field strengths and power confined in the low index gap region.

With the discovery of photonic crystals (cf. Sec. 9.1) another sort of optical waveguides appeared with a different guiding principle: Band gap guiding [12, 13, 60–62]. If defects are introduced in a photonic crystal with a complete band gap, they locally destroy the band gap properties and eigenstates with energies in the forbidden energy region can exist in the vicinity of those defects. At the same time, the light at such defects is locally confined, because the band gap is still intact in the surrounding.

This principle can be used to construct photonic crystal fibers (PCF), which are optical fibers with a two-dimensional photonic crystal structure in the cross section and a central defect as a core region. A typical PCF is displayed in Fig. 4.3. In this way, light is confined to the core region in the transverse

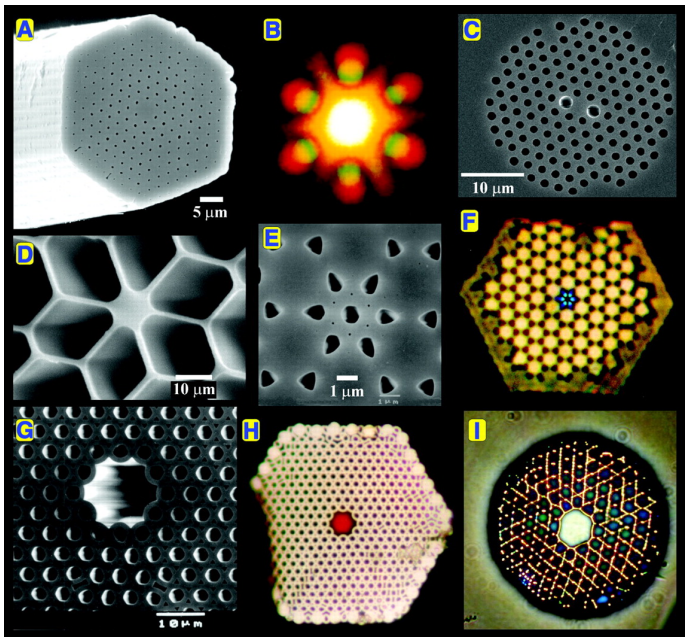


Figure 4.3.: *Selection of optical (OM) and scanning electron (SEM) micrographs of photonic crystal fibers (PCF) [12]. (A) SEM of an endlessly single-mode solid core PCF. (B) Far-field optical pattern produced by (A) when excited by red and green laser light. (C) SEM of a recent birefringent PCF. (D) SEM of a small (800 nm) core PCF with ultrahigh nonlinearity and a zero chromatic dispersion at 560-nm wavelength. (E) SEM of the first photonic band gap PCF, its core formed by an additional air hole in a graphite lattice of air holes. (F) Near-field OM of the six-leaved blue mode that appears when (E) is excited by white light. (G) SEM of a hollow-core photonic band gap fiber. (H) Near-field OM of a red mode in hollow-core PCF (white light is launched into the core). (I) OM of a hollow-core PCF with a Kagom cladding lattice, guiding white light. From [12]. Reprinted with permission from AAAS.*

plane and can be guided along the fiber axis. Typically used cross-sectional patterns for the cladding are hexagonally arranged air holes in a silica (SiO_2) background matrix. They exhibit complete band

gaps and are compatible to standard fiber drawing techniques. The defect that forms the core is then a completely left out air hole. Of course, there are many more different configurations which can be used. However, the design of a robust band gap is cumbersome and the technical production tolerances for hole diameters are tiny.

An interesting variation of the PCF is a *hollow core PCF*, also depicted in Fig. 4.3, where the defect is achieved by a much larger central air hole. It has the advantage that a great portion of the guided power is not confined to a dispersive medium but to dispersionless air.

Recent developments go in the direction of cheaper material systems like microstructured *polymer optical fibers* and other applications than telecommunications, e.g., optical sensing.[2]

Band gab guiding can also be achieved in three-dimensional photonic crystals, e.g., woodpile structures, either by incorporating continuous line defects [63–65], or by the evanescent coupling of separated point defects, called *coupled resonator optical waveguides* (CROW) [66].

4.1. Eigenmodes

There are three different types of eigenmodes of a waveguide structure: *Guided*, *leaky*, or *radiative* eigenmodes [57]. The most important for technical applications are the guided eigenmodes. They are characterized by a purely real propagation constant and transport electromagnetic energy without loss mainly within the core. In transverse direction their field strength decays exponentially and fast. However, in practice, lossless transport never occurs since there is either material absorption or scattering loss due to surface roughness, nonuniformities of the core cross section, or waveguide bends. Guided modes are the modes that remain a long distance away from the source — the spatially steady state. They have an effective refractive index above the guiding cutoff.

Leaky modes are characterized by a complex propagation constant with a rather small imaginary part. Despite a considerable amount of field within the core that resembles the field distribution of guided modes, they have oscillatory componentes outside the core. This means, they leak field and energy along the propagation and their amplitude decays. However, the modes usually travel a quite reasonable distances. Their effective refractive index is below the guiding cutoff, but close by.

Way below the guiding cutoff is the realm of an infinite number of radiative eigenmodes. A single radiative eigenmode has no proper physical meaning as their field is oscillatory and extends to infinity. Hence, they carry an unlimited amount of energy. Radiative modes get their physical meaning only as an entirety, as a superposition which ensures normalization and finite energy transport. These modes are essential to model scattering processes as well as fields in the spatially transient state near light sources.

Waveguides can occur either as *single-mode* or *multi-mode* waveguides, depending on the number of supported guided eigenmodes. Whether a waveguide is single- or multi-moded bases on the materials, the geometry, and the selected wavelength of the light. In general, the larger the geometry compared to the wavelength or the higher the core refractive index compared to the cladding, the more guided modes are supported.

[2] I had the great opportunity to get a profound insight in the development and production of polymer fibers whilst my stay at DTU.

4.1.1. Analytical Propagating Eigenmodes of Circular Step-Index Fibers

In this section we briefly introduce the analytical guided eigensolutions of a circular step-index fiber. We closely follow the derivations in Ref. [57], Chap. 12, and, therefore, only restate the most important steps. The circular step index fiber is one of the few systems where analytical solutions exist. Therefore, it is perfectly suited to serve as reference solution to benchmark our numerical results later on (cf. Chap. 8).

For the description of a circular step-index fiber we introduce the cylindrical polar coordinate system (r, ϕ, z), where the axis of the fiber with uniform cross section is assumed to coincide with the z-axis of the coordinate system. Then the fiber is fully described by its spatial permittivity distribution in the r-ϕ-plane

$$\varepsilon(\mathbf{r}, \omega) = \begin{cases} \varepsilon_{co}(\omega) & \text{for} \quad 0 \leq r < \rho\,, \\ \varepsilon_{cl}(\omega) & \text{for} \quad \rho < r < \infty\,, \end{cases} \tag{4.2}$$

with ρ the radius of the core and the spatially constant isotropic permittivity of the core ε_{co} and the cladding ε_{cl}. The permittivity at optical frequencies is $\mu(\mathbf{r}, \omega) = 1$. An appropriate separation ansatz for the electromagnetic modal fields is given by

$$\mathbf{E}(r, \phi, z) = (\mathbf{e}_t(r, \phi) + e_z(r, \phi)\,\hat{\mathbf{z}})\, e^{\mathrm{i}\beta z}\,, \qquad \mathbf{H}(r, \phi, z) = (\mathbf{h}_t(r, \phi) + h_z(r, \phi)\,\hat{\mathbf{z}})\, e^{\mathrm{i}\beta z}\,, \tag{4.3}$$

where we decomposed the fields into transversal components, labeled by subscript "t", and longitudinal components along z, and introduced the propagation constant β. This ansatz is substituited into the source free vector wave equations, Eqs. (2.24), and the differential operators are rewritten in cylindrical coordinates. Then, apart from the core-cladding interface within the core and cladding regions, the derivatives of the permittivity vanish and the electric and magnetic field components denoted by Ψ must obey

$$\left[\frac{\partial^2}{\partial R^2} + \frac{1}{R}\frac{\partial}{\partial R} + \frac{1}{R^2}\frac{\partial^2}{\partial \phi^2} + U^2\right] \Psi = 0\,, \qquad 0 \leq R < 1\,, \tag{4.4a}$$

$$\left[\frac{\partial^2}{\partial R^2} + \frac{1}{R}\frac{\partial}{\partial R} + \frac{1}{R^2}\frac{\partial^2}{\partial \phi^2} - W^2\right] \Psi = 0\,, \qquad 1 < R < \infty\,. \tag{4.4b}$$

Here, we introduce the normalized radius $R = r/\rho$ and we switch to dimensionless units as described in Sec. 2.1.5 with normalization constant $a = \rho$. Other parameters we introduce are the *core parameter*

$$U = \sqrt{\omega^2 \varepsilon_{co} - \beta^2} \tag{4.5}$$

and the *cladding parameter*

$$W = \sqrt{\beta^2 - \omega^2 \varepsilon_{cl}}\,. \tag{4.6}$$

Both parameters are connected by the *waveguide parameter*

$$V^2 = \omega^2(\varepsilon_{co} - \varepsilon_{cl}) = U^2 + W^2\,, \tag{4.7}$$

which solely depends on the permittivities of core and cladding, and the frequency of the electromagnetic fields.

It is sufficient to solve Eqs. (4.4) for the e_z and h_z components only, because the transverse components follow from

$$\mathbf{e}_t = \frac{\mathrm{i}}{P}\Big[\beta\,\boldsymbol{\nabla}_t e_z - \omega^2\,\hat{\mathbf{z}}\times\boldsymbol{\nabla}_t h_z\Big]\,, \tag{4.8a}$$

$$\mathbf{h}_t = \frac{\mathrm{i}}{P}\Big[\beta\,\boldsymbol{\nabla}_t h_z + \varepsilon\,\hat{\mathbf{z}}\times\boldsymbol{\nabla}_t e_z\Big] \tag{4.8b}$$

(cf. Ref. [57], Section 30-3), with P either U or W depending whether the transversal fields are considered in the core or cladding region, respectively. The transverse differential operator is given by $\boldsymbol{\nabla}_t\Psi = \hat{\mathbf{r}}\,\partial_r\Psi + \hat{\boldsymbol{\Phi}}\,\partial_\phi\Psi/r$.

If we take a closer look, we notice that the angular dependence in Eqs. (4.4) is fulfilled by sine or cosine terms with arguments $\nu\phi + \phi_0$ with $\nu \in \mathbb{N}_0$ and an arbitrary offset ϕ_0. Then the remaining radial part of Eq. (4.4a) in the core region is a Bessel differential equation in the argument UR with Bessel functions $J_\nu(UR)$ as bounded solutions, while the radial part of Eq. (4.4b) in the cladding region is a modified Bessel differential equation in the argument WR with modified Bessel functions of the second kind $K_\nu(WR)$ as bounded solutions accordingly. Hence, the longitudinal components of the fields must be of the general form

$$e_z = A\frac{J_\nu(UR)}{J_\nu(U)}\,f_\nu(\phi)\,, \qquad h_z = B\frac{J_\nu(UR)}{J_\nu(U)}\,g_\nu(\phi)\,, \qquad 0 \le R < 1\,, \tag{4.9a}$$

$$e_z = A\frac{K_\nu(WR)}{K_\nu(W)}\,f_\nu(\phi)\,, \qquad h_z = B\frac{K_\nu(WR)}{K_\nu(W)}\,g_\nu(\phi)\,, \qquad 1 < R < \infty\,. \tag{4.9b}$$

The denominators are normalization constants that guarantee the field continuity across the core / cladding interface. A and B are constants whose ratio can be fixed by the continuity condition of the azimuthal field component at $R = 1$. The angular dependence is consistently provided by

$$f_\nu(\phi) = \begin{cases} \cos\nu\phi \\ \sin\nu\phi \end{cases}, \qquad g_\nu(\phi) = \begin{cases} -\sin\nu\phi \\ \cos\nu\phi \end{cases}, \qquad \begin{matrix} \text{even modes} \\ \text{odd modes} \end{matrix}\,. \tag{4.10}$$

The solutions are, thus, characterized by ν full oscillation periods in ϕ-direction. The distinction between even and odd modes is accomplished by the choice of ϕ_0 with difference $\pi/2$ between the two polarizations such that any angular polarization can be constructed from the two orthogonal solutions by superposition.

There are three different non-trivial possibilities that Eqs. (4.4) are fulfilled: $e_z = 0$ and $h_z \neq 0$, $e_z \neq 0$ and $h_z = 0$, or $e_z \neq 0$ and $h_z \neq 0$. They are called *transverse electric* (TE) modes, *transverse magnetic* (TM) modes, and hybrid (HE and EH) modes, respectively. In order to fulfill the continuity conditions everywhere at the interface, the azimuthal parameter ν must be zero for both TE and TM modes and so must the derivatives ∂_ϕ. The resulting electromagnetic field distributions are summarized in Tab. 4.1. Here, we used the *profile height parameter* defined as $\Delta = (\varepsilon_{co} - \varepsilon_{cl})/2\varepsilon_{co}$.

However, there is still one parameter undetermined: the propagation constant β. From the remaining

(a) $\mathrm{HE}_{\nu m}$ and $\mathrm{EH}_{\nu m}$ modes

Component	Core	Cladding
e_r	$-\frac{a_1\,J_{\nu-1}(UR)+a_2\,J_{\nu+1}(UR)}{J_\nu(U)}\,f_\nu(\phi)$	$-\frac{U}{W}\,\frac{a_1\,K_{\nu-1}(WR)-a_2\,K_{\nu+1}(WR)}{K_\nu(W)}\,f_\nu(\phi)$
e_ϕ	$-\frac{a_1\,J_{\nu-1}(UR)-a_2\,J_{\nu+1}(UR)}{J_\nu(U)}\,g_\nu(\phi)$	$-\frac{U}{W}\,\frac{a_1\,K_{\nu-1}(WR)+a_2\,K_{\nu+1}(WR)}{K_\nu(W)}\,g_\nu(\phi)$
e_z	$-\frac{\mathrm{i}U}{\beta}\,\frac{J_\nu(UR)}{J_\nu(U)}\,f_\nu(\phi)$	$-\frac{\mathrm{i}U}{\beta}\,\frac{K_\nu(WR)}{K_\nu(W)}\,f_\nu(\phi)$
h_r	$-\frac{\varepsilon_{co}}{\beta}\,\frac{a_3\,J_{\nu-1}(UR)-a_4\,J_{\nu+1}(UR)}{J_\nu(U)}\,g_\nu(\phi)$	$-\frac{\varepsilon_{co}}{\beta}\,\frac{U}{W}\,\frac{a_5\,K_{\nu-1}(WR)+a_6\,K_{\nu+1}(WR)}{K_\nu(W)}\,g_\nu(\phi)$
h_ϕ	$-\frac{\varepsilon_{co}}{\beta}\,\frac{a_3\,J_{\nu-1}(UR)+a_4\,J_{\nu+1}(UR)}{J_\nu(U)}\,f_\nu(\phi)$	$-\frac{\varepsilon_{co}}{\beta}\,\frac{U}{W}\,\frac{a_5\,K_{\nu-1}(WR)-a_6\,K_{\nu+1}(WR)}{K_\nu(W)}\,f_\nu(\phi)$
h_z	$-\frac{\mathrm{i}UF_2}{\omega^2}\,\frac{J_\nu(UR)}{J_\nu(U)}\,g_\nu(\phi)$	$-\frac{\mathrm{i}UF_2}{\omega^2}\,\frac{K_\nu(WR)}{K_\nu(W)}\,g_\nu(\phi)$

$a_1=\frac{F_2-1}{2}$ $\quad a_3=\frac{F_1-1}{2}$ $\quad a_5=\frac{F_1-1+2\Delta}{2}$ $\quad F_1=\left(\frac{UW}{V}\right)^2\frac{b_1+(1-2\Delta)b_2}{\nu}$ $\quad F_2=\left(\frac{V}{UW}\right)^2\frac{\nu}{b_1+b_2}$

$a_2=\frac{F_2+1}{2}$ $\quad a_4=\frac{F_1+1}{2}$ $\quad a_6=\frac{F_1+1-2\Delta}{2}$ $\quad b_1=\frac{1}{2U}\left(\frac{J_{\nu-1}(U)}{J_\nu(U)}-\frac{J_{\nu+1}(U)}{J_\nu(U)}\right)$ $\quad b_2=\frac{-1}{2W}\left(\frac{K_{\nu-1}(W)}{k_\nu(w)}+\frac{K_{\nu+1}(W)}{K_\nu(W)}\right)$

(b) TE_{0m} modes

Component	Core	Cladding
e_ϕ	$-\frac{J_1(UR)}{J_1(U)}$	$-\frac{K_1(WR)}{K_1(W)}$
h_r	$\frac{\beta}{\omega^2}\,\frac{J_1(UR)}{J_1(U)}$	$\frac{\beta}{\omega^2}\,\frac{K_1(WR)}{K_1(W)}$
h_z	$\frac{\mathrm{i}U}{\omega^2}\,\frac{J_0(UR)}{J_1(U)}$	$\frac{-\mathrm{i}U}{\omega^2}\,\frac{K_0(WR)}{K_1(W)}$
$e_r=e_z=h_\phi=0$		

(c) TM_{0m} modes

Component	Core	Cladding
e_r	$\frac{J_1(UR)}{J_1(U)}$	$\frac{\varepsilon_{co}}{\varepsilon_{cl}}\,\frac{K_1(WR)}{K_1(W)}$
e_z	$\frac{\mathrm{i}U}{\beta}\,\frac{J_0(UR)}{J_1(U)}$	$\frac{-\mathrm{i}W}{\beta}\,\frac{\varepsilon_{co}}{\varepsilon_{cl}}\,\frac{K_0(WR)}{K_1(W)}$
h_ϕ	$\frac{\varepsilon_{co}}{\beta}\,\frac{J_1(UR)}{J_1(U)}$	$\frac{\varepsilon_{co}}{\beta}\,\frac{K_1(WR)}{K_1(W)}$
$e_\phi=h_r=h_z=0$		

Table 4.1.: *Eigenmode field components of the circular step-index fiber in dimensionless units [57].*

continuity conditions we can derive the eigenvalue equations [57, 67]

$$\mathrm{HE}_{\nu m}\text{ and }\mathrm{EH}_{\nu m}: \quad \left[\frac{J'_\nu(U)}{UJ_\nu(U)}+\frac{K'_\nu(W)}{WK_\nu(W)}\right]\left[\frac{J'_\nu(U)}{UJ_\nu(U)}+\frac{\varepsilon_{cl}}{\varepsilon_{co}}\frac{K'_\nu(W)}{WK_\nu(W)}\right] -\nu^2\left(\frac{1}{U^2}+\frac{1}{W^2}\right)\left(\frac{1}{U^2}+\frac{\varepsilon_{cl}}{\varepsilon_{co}}\frac{1}{W^2}\right)=0\,, \tag{4.11a}$$

$$\mathrm{TE}_{0m}: \quad \left[\frac{J'_1(U)}{UJ_0(U)}+\frac{K'_1(W)}{WK_0(W)}\right]=0\,, \tag{4.11b}$$

$$\mathrm{TM}_{0m}: \quad \left[\frac{J'_1(U)}{UJ_0(U)}+\frac{\varepsilon_{cl}}{\varepsilon_{co}}\frac{K'_1(W)}{WK_0(W)}\right]=0\,. \tag{4.11c}$$

The index m denotes the m-th root of the eigenvalue equation starting from the largest β value. These equations are transcendental and have to be solved numerically.

For the numerical solution of the eigenvalue equations we use a self-developed MATLAB code. The task is to calculate the propagation constant $\beta_{\nu m}$ from the m-th root $U_{\nu m}$ of Eqs. (4.11) for given ν. The parameters ε_{co}, ε_{cl}, and V are fixed by the geometry and permittivity profile of the fiber. Then, the cladding parameter $W = W(U, V)$ is a function of U only. The core parameter can take values $0 < U \leq V$. The core parameter and, thus, the propagation constant are monotonically increasing functions of V.

The roots $U_{\nu m}$ are determined using the built-in MATLAB function *fzero*, which takes as input parameter a starting value that should be as close as possible to the root value where the eigenvalue equation has a sign change. Hence, we must estimate a proper starting value.

This is a rather easy task for TE and TM modes since the values the root can take are restricted to a finite interval – called root interval – whose lower boundary is determined by the mode cutoff, where $U = V$ and $W = 0$, and whose upper boundary is obtained from the limit $V \to \infty$. Substituting these conditions into the eigenvalue equations leads to the upper and lower interval boundaries calculated from the zeros of the corresponding (Bessel) functions noted down in Tab. 4.2 [57]. The

Mode	Mode cutoff ($U \to V$, $W \to 0$)	Upper limit ($V \to \infty$, $W \to \infty$)
TE_{0m}, TM_{0m}	$J_0(U) = 0$	$J_1(U) = 0$
HE_{1m}	$J_1(U) = 0$	$J_0(U) = 0$
$\mathrm{HE}_{\nu m}$ ($\nu > 1$)	$\frac{U}{\nu-1}\frac{J_{\nu-2}(U)}{J_{\nu-1}(U)} + \frac{2\Delta}{1-2\Delta} = 0$	$J_{\nu-1}(U) = 0$
$\mathrm{EH}_{\nu m}$	$J_\nu(U) = 0$	$J_{\nu+1}(U) = 0$

Table 4.2.: *Conditions for the lower (left column) and upper (right column) boundaries of the root intervals for $U_{\nu m}$ [57].*

necessary roots of the Bessel functions are calculated using the public domain function *zerobess* from MATLAB Central File Exchange written by J. Lundgren [68]. For TE and TM solutions we use the center of the obtained intervals as starting value.

The procedure for hybrid modes is more tricky. The curves given by Eqs. (4.11) often only slightly extends into the negative region, such that two roots are very close together. The roots where the function changes from positive to negative values with increasing V correspond to EH modes. Where it changes from negative to positive values, the root belongs to HE modes. The intervals of values which $U_{\nu m}$ can reach usually overlap; they are marked in the figure as well. Consequently, a poorly chosen starting value can lead to either of the roots. To make this procedure more reliable we pick the starting value for EH modes not in the center of the interval as before, but ten percent of the interval size to the upper side of the lower bound and for HE modes ten percent of the interval size to the lower side of the upper bound. If the *fzero* function does not return a proper output, we successively move the starting value closer to the center of the interval and repeat the procedure. With this scheme we obtained satisfying results with a high reliability.

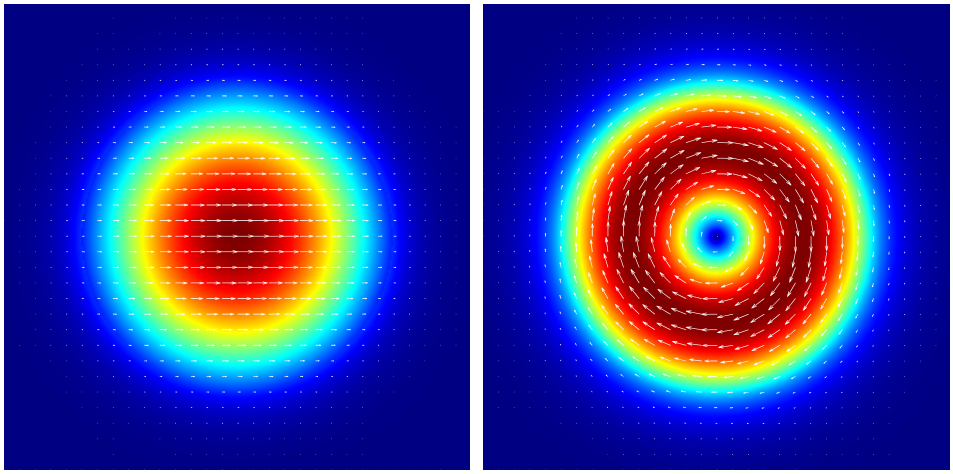
Next, we present the eigenmodes of a system we will investigate later on with our numerical simulation tools as an example (cf. Chap. 8). The analytical solutions will be our references the numerical solutions are compared against. The considered system is a step-index fiber as introduced above with a circular core of radius $\rho = 2.15\,\mu\mathrm{m}$ and permittivity $\varepsilon_{co} = 2.5$, and a uniform isotropic infinitely extended cladding with $\varepsilon_{cl} = 2.0952074$. For the examined free space wavelength $\lambda_{\mathrm{SI}} = 1.25\,\mu\mathrm{m}$

Mode	$U_{\nu m}$	$\mathrm{n_{eff}}$
HE_{11}	2.1178257366	1.5689477743
TE_{01}	3.3277325680	1.5508656652
HE_{21}	3.3667902790	1.5501437158
TM_{01}	3.3854981813	1.5497948115
EH_{11}	4.4677029833	1.5261376435
HE_{31}	4.4981666808	1.5253712642
HE_{12}	4.7990216858	1.5175006510

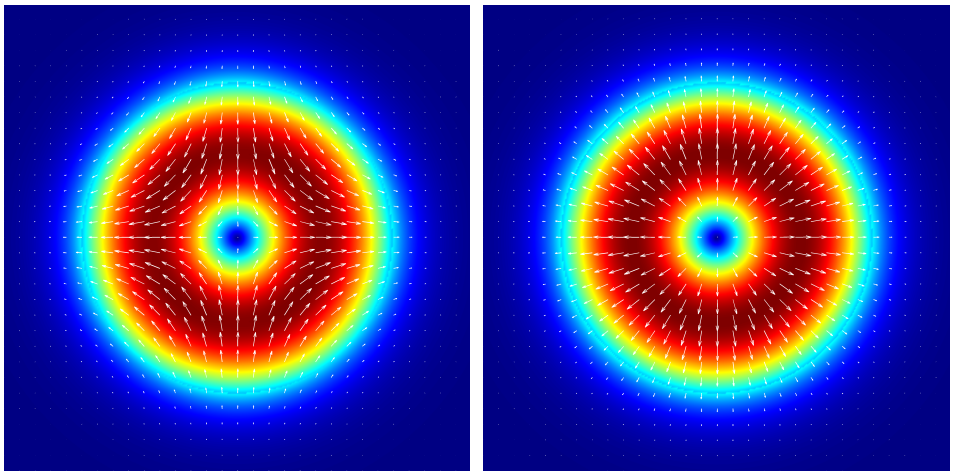
Mode	$U_{\nu m}$	$\mathrm{n_{eff}}$
EH_{21}	5.5114247070	1.4966353940
HE_{41}	5.5633662491	1.4949890269
TE_{02}	5.9691022751	1.4815291627
HE_{22}	6.0149955518	1.4799390397
TM_{02}	6.0222652226	1.4796858832
EH_{31}	6.4947167622	1.4624761683
HE_{51}	6.5728069662	1.4594859793

Table 4.3.: *Eigenmodes of the step-index fiber: Roots of the eigenvalue equations $U_{\nu m}$, and corresponding effective refractive indices $\mathrm{n_{eff}}$ in decreasing order.*

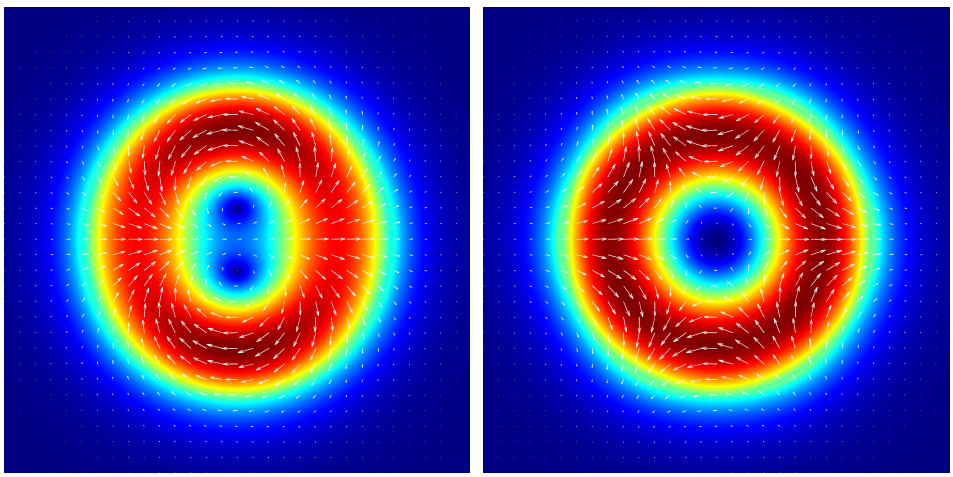
the waveguide parameter is fixed to $V = 6.875822$. The calculated effective refractive indices are gathered in Tab. 4.3. Plots of the transverse electric field $\mathbf{E}_t = (E_x, E_y)^T$ are shown in Fig. 4.4 separated into a color coded magnitude and white arrows indicating directions. Please note that only even hybrid modes (HE and EH) are depicted. The respective odd modes can be obtained through a rotation by $\pi/(2\nu)$. The total number of guided modes for this configuration then amounts to 24.



(a) HE_{11} **(b)** TE_{01}

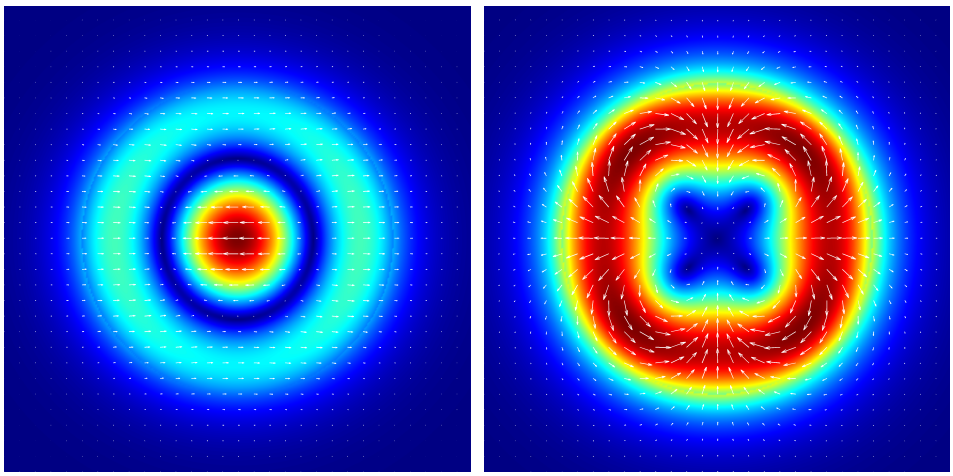


(c) HE_{21} **(d)** TM_{01}

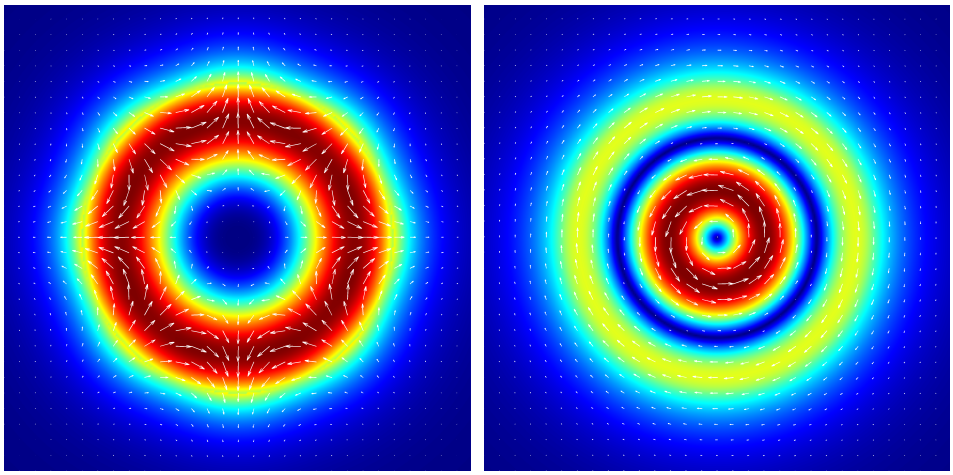


(e) EH_{11} **(f)** HE_{31}

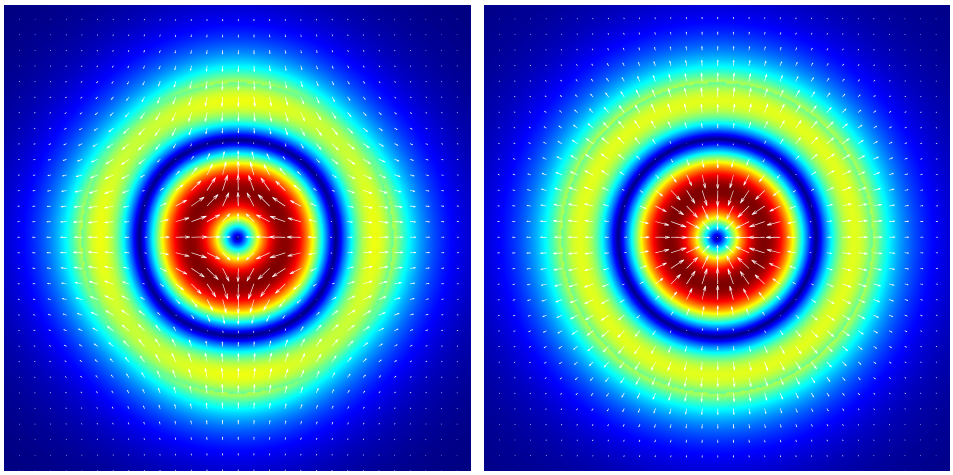
Figure 4.4.: *Eigenmodes 1 to 6 of a circular step-index fiber sorted by decreasing effective refractive indices. The color indicates the strength of the normalized transverse electric field (linear scale), the white arrows indicate its direction.*



(g) HE_{12} **(h)** EH_{21}

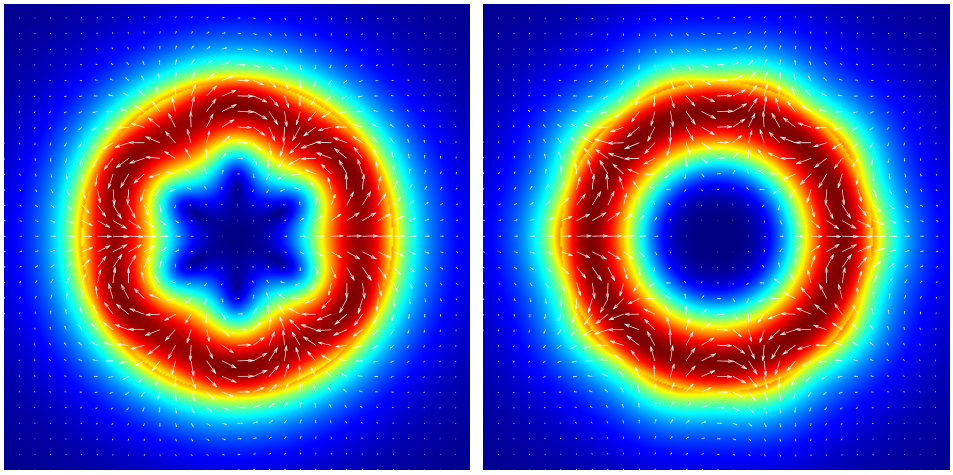


(i) HE_{41} **(j)** TE_{02}



(k) HE_{22} **(l)** TM_{02}

Figure 4.4.: *Eigenmodes 7 to 12 of a circular step-index fiber sorted by decreasing effective refractive indices. The color indicates the strength of the normalized transverse electric field (linear scale), the white arrows indicate its direction.*



(m) EH_{31} **(n)** HE_{51}

Figure 4.4.: *Eigenmodes 13 and 14 of a circular step-index fiber sorted by decreasing effective refractive indices. The color indicates the strength of the normalized transverse electric field (linear scale), the white arrows indicate its direction.*

4.1.2. Reciprocity Theorem and Mode Orthogonality

The aim of this section is to establish an *orthogonality equation* for the eigenmodes of a waveguide. An orthogonality condition is handy, for example, when we discuss the matching of fields in modal methods in Sec. 5.3.1. To this end, we start with the derivation of the *reciprocity theorem* as an important integral relationship between two solutions of Maxwell's equations. The reciprocity theorem is the basis for the proof and derivation of many modal quantities and properties. Besides the orthogonality, the relations for power flow, the relations for modal amplitudes due to current sources, as well as expressions for the group velocity follow from the reciprocity theorem [57], for instance.

Reciprocity Theorem

There are two forms of the reciprocity theorem — one with and one without complex conjugated fields. The former is valid for non-absorbing systems only. This is why we focus on the unconjugated form which is valid for non-absorbing as well as absorbing systems. For a detailed discussion of the reciprocity theorem please refer to Ref. [57].

The reciprocity theorem requires a vector function $\mathbf{V}$ defined as

$$\mathbf{V} = \mathbf{E} \times \bar{\mathbf{H}} + \bar{\mathbf{E}} \times \mathbf{H}\,, \tag{4.12}$$

where the unbarred and barred fields denote the solutions of two distinct (guiding) structures characterized by the respective material parameters $\underline{\varepsilon}$, $\underline{\mu}$ and $\bar{\underline{\varepsilon}}$, $\bar{\underline{\mu}}$ ($\mathbf{J} = \bar{\mathbf{J}} = 0$). The integral identity [57]

$$\int_A \boldsymbol{\nabla} \cdot \mathbf{V}\, \mathrm{d}A = \partial_z \int_A \mathbf{V} \cdot \hat{\mathbf{z}}\, \mathrm{d}A + \oint_{\partial A} \mathbf{V} \cdot \hat{\mathbf{n}}\, \mathrm{d}l \tag{4.13}$$

applies to a planar surface A orthogonal to the unit vector in propagation direction $\hat{\mathbf{z}}$ with perimeter ∂A, on which $\hat{\mathbf{n}}$ is the unit outward normal. The product $\mathbf{V} \cdot \hat{\mathbf{n}}$ only depends on the electric and magnetic fields normal to $\hat{\mathbf{n}}$ which are parallel to the interfaces of the waveguide, i.e., the core/cladding interface. According to the continuity conditions, Eqs. (2.19), these fields are continuous. Consequently, the line integral over the perimeter vanishes for all physical fields which must decay to zero at infinite distance to guarantee energy conservation. If we used periodic boundary conditions, the line integral would vanish as well, because the contributions from opposite sides of the unit cell (UC) boundary would cancel each other. Thus, we drop the line integral in Eq. (4.13) for an appropriate integration area A' ($A' = A_\infty$ or $A' = A_{\mathrm{UC}}$), and obtain the unconjugated form of the reciprocity theorem:

$$\partial_z \int_{A'} \mathbf{V} \cdot \hat{\mathbf{z}} \, \mathrm{d}A = \int_{A'} \boldsymbol{\nabla} \cdot \mathbf{V} \, \mathrm{d}A \,. \tag{4.14}$$

For completeness, we also introduce the conjugated reciprocity theorem. In the conjugated form of the theorem the vector function $\mathbf{V}$ is just replaced by

$$\mathbf{V}_c = \mathbf{E} \times \bar{\mathbf{H}}^* + \bar{\mathbf{E}}^* \times \mathbf{H} \,, \tag{4.15}$$

where the asterisk denotes the solutions of the complex conjugated Maxwell's equations.

Mode Orthogonality

Let us now examine the divergence term on the right hand side of Eq. (4.14) in more detail. Using the vector identity

$$\boldsymbol{\nabla} \cdot (\mathbf{E} \times \mathbf{H}) = \mathbf{H} \cdot (\boldsymbol{\nabla} \times \mathbf{E}) - \mathbf{E} \cdot (\boldsymbol{\nabla} \times \mathbf{H}) \,, \tag{4.16}$$

we can rewrite the divergence term into expressions that involve curls of the electric and magnetic fields which are familiar from Maxwell's equations. Since we are mainly interested in the orthogonality relation of eigenmodes, we work with the specialized form of source-free, time-harmonic Maxwell's equations as given in Eq. (2.13). Substituting these relations into Eq. (4.16), we deduce that the divergence term is given by

$$\boldsymbol{\nabla} \cdot \mathbf{V} = \mathrm{i}\omega^2 \left(\bar{\mathbf{H}} \underline{\boldsymbol{\mu}} \mathbf{H} - \mathbf{H} \underline{\bar{\boldsymbol{\mu}}} \bar{\mathbf{H}} \right) + \mathrm{i} \left(\mathbf{E} \underline{\bar{\boldsymbol{\varepsilon}}} \bar{\mathbf{E}} - \bar{\mathbf{E}} \underline{\boldsymbol{\varepsilon}} \mathbf{E} \right) . \tag{4.17}$$

Using general Maxwell's equations including sources would provide further terms containing the free current densities $\mathbf{J}$ and $\bar{\mathbf{J}}$. However, note that the divergence of $\mathbf{V}$, Eq. (4.17), vanishes if $\underline{\boldsymbol{\varepsilon}} = \underline{\bar{\boldsymbol{\varepsilon}}}^T$ and $\underline{\boldsymbol{\mu}} = \underline{\bar{\boldsymbol{\mu}}}^T$. Then, what remains from Eq. (4.14) is

$$\partial_z \int_{A'} \mathbf{V} \cdot \hat{\mathbf{z}} \, \mathrm{d}A = 0 \,. \tag{4.18}$$

We use this equation to derive the orthogonality condition for modes of the same waveguide.

Consider two forward traveling modes with propagation constants β_j and β_k. Their fields are given by

$$\mathbf{E}(x,y,z) = \mathbf{E}_j(x,y)\, e^{\mathrm{i}\beta_j z} \,, \qquad \mathbf{H}(x,y,z) = \mathbf{H}_j(x,y)\, e^{\mathrm{i}\beta_j z} \,, \tag{4.19a}$$

and

$$\bar{\mathbf{E}}(x,y,z) = \mathbf{E}_k(x,y)\, e^{\mathrm{i}\beta_k z} \,, \qquad \bar{\mathbf{H}}(x,y,z) = \mathbf{H}_k(x,y)\, e^{\mathrm{i}\beta_k z} \,. \tag{4.19b}$$

Substituting Eqs. (4.19) and Eq. (4.12) into Eq. (4.18) yields after some simple manipulation

$$\int_{A'} (\mathbf{E}_j \times \mathbf{H}_k + \mathbf{E}_k \times \mathbf{H}_j) \cdot \hat{\mathbf{z}} \, \mathrm{d}A = 0 \qquad \text{for } (\beta_j + \beta_k) \neq 0 \,. \tag{4.20}$$

Next, we focus on the same forward-traveling mode j and a backward-traveling mode with fields

$$\bar{\mathbf{E}}(x,y,z) = \mathbf{E}_{-k}(x,y)\, e^{\mathrm{i}\beta_{-k}z} \,, \qquad \bar{\mathbf{H}}(x,y,z) = \mathbf{H}_{-k}(x,y)\, e^{\mathrm{i}\beta_{-k}z} \,. \tag{4.21}$$

Substituting these fields into Eq. (4.18) is equivalent to replacing $\mathbf{E}_k \to \mathbf{E}_{-k}$, $\mathbf{H}_k \to \mathbf{H}_{-k}$, and $\beta_k \to \beta_{-k}$ in Eq. (4.20) which leads to

$$\int_{A'} (\mathbf{E}_j \times \mathbf{H}_{-k} + \mathbf{E}_{-k} \times \mathbf{H}_j) \cdot \hat{\mathbf{z}} \, \mathrm{d}A = 0 \qquad \text{for } (\beta_j + \beta_{-k}) \neq 0 \,. \tag{4.22}$$

A backward traveling mode ("-") and the corresponding forward traveling mode ("+") are related by [57]

$$\mathbf{E}^- = \quad \mathbf{E}_t^+ - E_z^+ \hat{\mathbf{z}} \,, \tag{4.23a}$$
$$\mathbf{H}^- = -\mathbf{H}_t^+ + H_z^+ \hat{\mathbf{z}} \,, \tag{4.23b}$$
$$\beta_{-k} = -\beta_k \,, \tag{4.23c}$$

where subscript "t" denotes the components normal to the propagation direction.[3] With this relationship we can express the vector products of forward modes by those involving backward modes and vice versa. This means we substitute

$$(\mathbf{E}_j \times \mathbf{H}_k) \cdot \hat{\mathbf{z}} = -\,(\mathbf{E}_j \times \mathbf{H}_{-k}) \cdot \hat{\mathbf{z}} \,, \tag{4.24a}$$
$$(\mathbf{E}_k \times \mathbf{H}_j) \cdot \hat{\mathbf{z}} = \quad (\mathbf{E}_{-k} \times \mathbf{H}_j) \cdot \hat{\mathbf{z}} \,, \tag{4.24b}$$

which even holds for modes from different waveguides, and Eq. (4.23c) into Eq. (4.22). We obtain

$$\int_{A'} (-\mathbf{E}_j \times \mathbf{H}_k + \mathbf{E}_k \times \mathbf{H}_j) \cdot \hat{\mathbf{z}} \, \mathrm{d}A = 0 \qquad \text{for } (\beta_j - \beta_k) \neq 0 \,. \tag{4.25}$$

Addition and subtraction of Eq. (4.20) and Eq. (4.25) finally provides the orthogonality equation for two forward traveling modes

$$\int_{A'} (\mathbf{E}_j \times \mathbf{H}_k) \cdot \hat{\mathbf{z}} \, \mathrm{d}A = \int_{A'} (\mathbf{E}_k \times \mathbf{H}_j) \cdot \hat{\mathbf{z}} \, \mathrm{d}A = 0 \qquad \text{for } (\beta_j \pm \beta_k) \neq 0 \,. \tag{4.26a}$$

Taking Eqs. (4.24) into account, an equivalent relation for a forward and a backward traveling mode is found:

$$\int_{A'} (\mathbf{E}_j \times \mathbf{H}_{-k}) \cdot \hat{\mathbf{z}} \, \mathrm{d}A = \int_{A'} (\mathbf{E}_{-k} \times \mathbf{H}_j) \cdot \hat{\mathbf{z}} \, \mathrm{d}A = 0 \qquad \text{for } (\beta_j \pm \beta_{-k}) \neq 0 \,. \tag{4.26b}$$

In absorbing waveguides, these equations even hold for *leaky modes* with complex propagation constants which usually occur there. Furthermore, all modes are also orthogonal to the total radiation field $\bar{\mathbf{E}} = \mathbf{E}_{\mathrm{rad}}$, $\bar{\mathbf{H}} = \mathbf{H}_{\mathrm{rad}}$ built up by a superposition of all excited *radiation modes* [57].

[3] There are actually two possible conventions which solve Maxwell's equations consistently [57]. The second is:
$\mathbf{E}^- = -\mathbf{E}_t^+ + E_z^+ \hat{\mathbf{z}} \,, \qquad \mathbf{H}^- = \mathbf{H}_t^+ - H_z^+ \hat{\mathbf{z}} \,.$
We chose the convention of Eq. (4.23).

Chapter 5.

5 Modal Methods

The idea of modal methods is the description of electromagnetic fields in terms of eigenmodes of the investigated structure. When we denote the electromagnetic fields with $\mathbf{\Psi}(\mathbf{r})$, where $\mathbf{\Psi}$ comprises electric and magnetic field components, and the structure's eigenmodes with $\mathbf{\Psi}_j(\mathbf{r})$, the field expansion is given by

$$\mathbf{\Psi}(\mathbf{r}) = \sum_j a_j \, \mathbf{\Psi}_j(\mathbf{r}) \, . \tag{5.1}$$

It does not matter whether the eigenmodes are those of periodic systems like Bloch modes (cf. Sec. 3.3), or those of non-periodic systems like eigenmodes of a waveguide as introduced in Chap. 4.

In most cases we are interested in the propagation of electromagnetic fields in a distinguishable direction. The advantage of modal methods is the easy and efficient description of propagation in this primary direction. In a uniform waveguide, for example, the primary direction is the direction along its axis. The propagation along this homogeneous direction is entirely described by a plane wave with appropriate propagation constant. Similarly, in any periodic structure the propagation is determined by the Bloch phase.

Real world structures are of finite size and, as such, often dominated by boundary effects which do not easily comply with Bloch periodic eigenmodes. Furthermore, with modal methods we would like to be able to handle structures with a variety of shapes as broad as possible, deviating very much from perfect periodicity. From this point of view, it seems wise to prefer waveguide-like eigenmodes of structures homogeneous along the primary direction over eigenmodes of structures periodic along the primary direction. Nevertheless, there are situations where the latter are to be favored [69]. Still, for the course of this work we focus on the former. Hence, any arbitrary structure we would like to investigate must first be homogenized along the primary direction.

We establish the convention that the primary direction coincides with the z-axis of the Cartesian coordinate system. Furthermore, from here on, we solely use the covariant notation valid in general curvilinear coordinate systems as introduced in Sec. 2.5. This means that for the remainder of this work our curvilinear coordinate system $\mathcal{O}x^1x^2x^3$ is chosen such that the $\mathbf{e}_3$-direction is fixed to always coincide with our primary direction along the Cartesian z-axis. The directions $\mathbf{e}_1$ and $\mathbf{e}_2$ still remain free, but restricted to the plane orthogonal to $\mathbf{e}_3$.[1] Then x^1 and x^2 ($\mathbf{e}_1$ and $\mathbf{e}_2$) can be referred to as *transverse* coordinates (directions).

[1] This implies that $\mathbf{e}_1$ and $\mathbf{e}_2$ need not be orthogonal to each other.

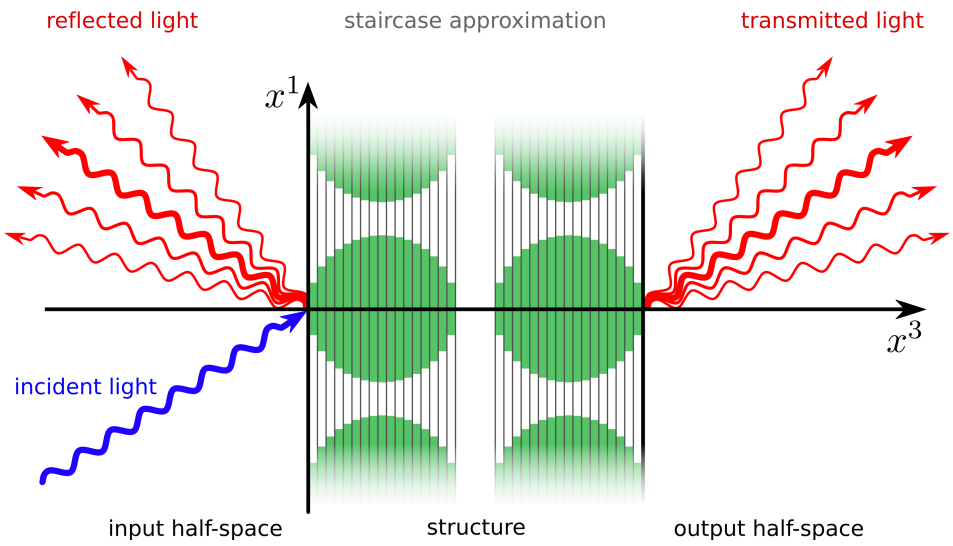


Figure 5.1.: *Schematic illustration of a staircase approximated system. The cylindrical structures are piecewise homogenized along the main propagation direction x^3. The result looks like a staircase and decomposes into several layers.*

The piecewise homogenization along the primary direction (x^3-direction) is called staircase approximation. A generic example which visualizes this procedure is shown in Fig. 5.1. In this way, every three-dimensional structure is decomposed into a set of independent layers.

The homogeneity of a layer in the primary direction allows for a plane wave ansatz

$$\mathbf{E}(x^1, x^2, x^3) = \mathbf{E}(x^1, x^2)\, e^{\mathrm{i}k_3 x^3} \qquad \text{and} \qquad \mathbf{H}(x^1, x^2, x^3) = \mathbf{H}(x^1, x^2)\, e^{\mathrm{i}k_3 x^3} \tag{5.2}$$

for the fields. This ansatz transforms each layer into an effectively two-dimensional subsystem (x^1-x^2-plane) to be solved for its specific set of eigenmodes. The eigenmodes and the corresponding propagation constants k_3 need to be calculated from the eigenvalue problem that can be derived from Maxwell's equations. With the eigenmodes at hand, the fields in each layer are expanded into the eigenmode basis. This ensures an easy determination of the field evolution in x^3-direction.

In order to obtain the solution for the entire structure, the subsystems are connected by the continuity of the fields at the interface between adjacent layers. This is formalized within the scattering matrix (S-matrix) algorithm. The scattering matrix consistently connects the expansion amplitudes of incident electromagnetic waves to the amplitudes of outgoing waves in the first and last layers. If the incident waves are given, the scattering matrix provides the reflected and transmitted fields, and, with some further effort, also the fields inside the structure.

Here, we give a short overview of the procedure:

1. The structure is decomposed by a staircase-approximation into a set of layers which are each homogeneous in x^3-direction .
2. In each layer the eigenvalue problem is solved and the eigenmodes and their propagation constants are calculated.
3. The fields are expanded into the layer's specific eigenmode basis.

4. The scattering matrix for the entire structure is constructed by demanding continuity of the tangential fields at the interfaces between adjacent layers.
5. For incoming electromagnetic waves the scattering matrix provides the reflected and transmitted fields.

The primary difference between different modal methods is the way of solving the eigenvalue equation. Simple structures like radially symmetric waveguides can be solved analytically, whereas complex structures can usually only be solved numerically. The various numerical methods differ in their spatial discretization and the choice of the basis functions. Finite element methods (FEM) [70], for example, usually work with unstructured grids and polynomial basis functions[2], while the B-spline modal method (BMM) [53, 73, 74] utilizes structured meshes and special piecewise polynomial basis functions, the basis-splines (B-splines. Finally, the Fourier modal method (FMM) [24, 75] works with structured grids and plane waves as basis functions. Each combination has its advantages and disadvantages.

Before we get to a specific method and the discretization of the problem, we present the mentioned common aspects, equations and algorithms of modal methods in more detail in the subsequent sections. The discussion is roughly oriented along the beautiful abstracted scheme of modal methods established in Ref. [53]. The structure's decomposition is covered in Sec. 5.1. Section 5.2 is concerned with the formulation of the eigenproblem and the expansion of the fields in eigenmodes. The S-matrix algorithm is derived and introduced in different variations in Sec. 5.3. The method of choice in this work is the FMM which will be presented together with all method specific aspects in Chap. 6.

5.1. Structure Decomposition into Layers

The decomposition of the structure under investigation into L layers by a staircase approximation is illustrated in Fig. 5.2. When we approximate a structure, the layers do not have to be of equal thickness $t^{(l)} = (x_l^3 - x_{l-1}^3)$, though in most cases they are. Subscript or bracketed superscript labels $l = 1, \ldots, L$ indicate the associated layer. The slicing is usually done by defining the layer centers $\bar{z}^{(l)} = x_{l-1}^3 + d^{(l)}/2$. Layer 1 is the input layer, and layer L the output layer. The l-th layer is characterized by its specific permittivity

$$\underline{\boldsymbol{\varepsilon}}^{(l)}(x^1, x^2) = \underline{\boldsymbol{\varepsilon}}(x^1, x^2, \bar{z}^{(l)})\,, \tag{5.3a}$$

and its specific permeability

$$\underline{\boldsymbol{\mu}}^{(l)}(x^1, x^2) = \underline{\boldsymbol{\mu}}(x^1, x^2, \bar{z}^{(l)})\,. \tag{5.3b}$$

Different materials in Fig. 5.2 are sketched by colors (green, white).

In Sec. 5.2 we focus on the eigenmodes of a single layer only. Therefore, the layer labels will be omitted.

[2] In electromagnetic applications, the basis functions are usually vector polynomials [71, 72].

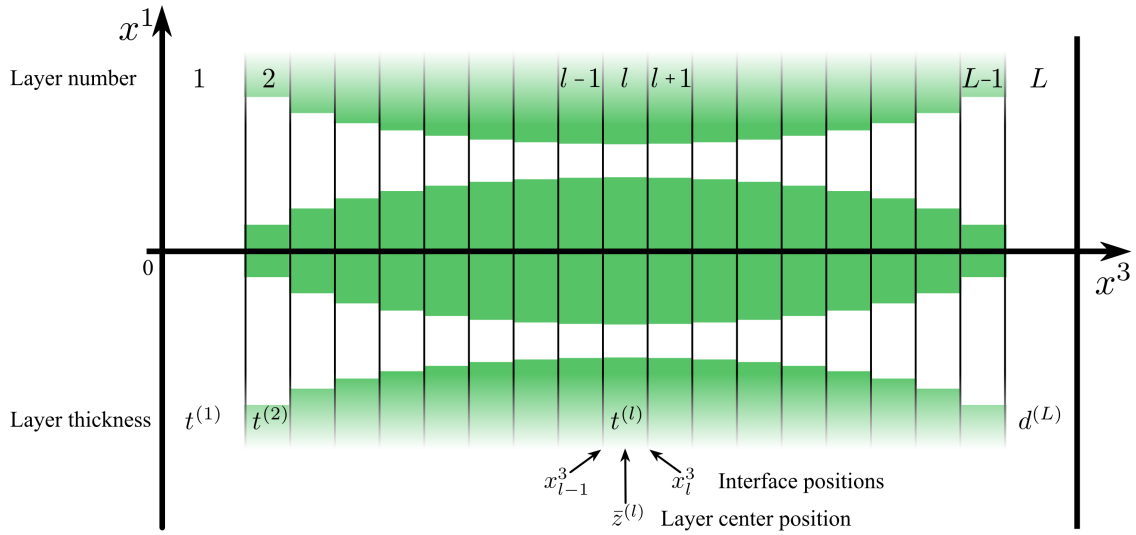


Figure 5.2.: *Schematic illustration of layer numbering l, layer thicknesses $t^{(l)}$, interface positions x^3_l, and layer center positions $\bar{z}^{(l)}$. For further details see text.*

5.2. Eigenmode Expansion

The eigenmode expansion is fundamental to modal methods. Before we can do the expansion, the eigenmodes have to be calculated first. Since structure and material parameters differ from layer to layer, the eigenmodes are specific to a layer. More precisely, only if the structure geometry or the material parameters vary, the eigenmodes are different. However, they are the same if the layers can be distinguished only by their variation in thickness. The eigenmodes can be calculated as solutions of an eigenvalue problem which is derived from Maxwell's equations using the plane wave ansatz for the x^3-dependence. The eigenvalue equation we describe, must be solved for each distinguishable layer separately.

5.2.1. Derivation of the Eigenvalue Problem

We start with the dimensionless Maxwell's curl equations in frequency domain and in covariant notation as given in Eq. (2.84). It is convenient to restate them here:

$$\epsilon^{\rho\sigma\tau}\,\frac{\partial_\sigma}{\mathrm{i}}\,E_\tau = \omega^2\,\mu^{\rho\sigma}\,H_\sigma\,, \tag{5.4a}$$

$$\epsilon^{\rho\sigma\tau}\,\frac{\partial_\sigma}{\mathrm{i}}\,H_\tau = -\,\varepsilon^{\rho\sigma}\,E_\sigma\,. \tag{5.4b}$$

Equations 5.4 include the full anisotropy of permittivity and permeability with nine tensor components each, and thus are the most general form of Maxwell's equations. From them, we derive the full anisotropic eigenproblem. All special cases can easily be deduced by setting some components of the material quantities to zero. Note that the full derivation can be found in App. B.1 – we only state the major steps and results here.

We solve the third components of Eq. (5.4a) and Eq. (5.4b) for the field components H_3 and E_3,

respectively, and obtain[3]

$$H_3 = \frac{1}{\omega^2}\,\breve{\mu}^{33}\left(-\frac{\partial_1}{\mathrm{i}}\left(-E_2\right) - \frac{\partial_2}{\mathrm{i}}E_1\right) - \breve{\mu}^{31}H_1 - \breve{\mu}^{32}H_2\,, \tag{5.5a}$$

$$E_3 = \breve{\varepsilon}^{33}\left(\frac{\partial_2}{\mathrm{i}}H_1 - \frac{\partial_1}{\mathrm{i}}H_2\right) - \breve{\varepsilon}^{31}E_1 + \breve{\varepsilon}^{32}\left(-E_2\right)\,. \tag{5.5b}$$

Here, we introduce new quantities for the permittivity $\underline{\breve{\boldsymbol{\varepsilon}}}$ and permeability $\underline{\breve{\boldsymbol{\mu}}}$ which contain aggregates of components of the old material parameters $\underline{\boldsymbol{\varepsilon}}$ and $\underline{\boldsymbol{\mu}}$. They are defined by the $\hat{\mathrm{l}}_3^-$ operator of Ref. [51] (cf. Sec. 3.4.4). The operator takes care that the terms in the equations are grouped according to the different transversal field components E_1, $\left(-E_2\right)$, H_1, and H_2. The components of the new permittivity $\underline{\breve{\boldsymbol{\varepsilon}}} = \hat{\mathrm{l}}_3^-\left(\underline{\boldsymbol{\varepsilon}}\right)$ are given by

$$\begin{aligned}
\breve{\varepsilon}^{11} &= \varepsilon^{11} - \varepsilon^{13}(\varepsilon^{33})^{-1}\varepsilon^{31}\,,\\
\breve{\varepsilon}^{12} &= \varepsilon^{12} - \varepsilon^{13}(\varepsilon^{33})^{-1}\varepsilon^{32}\,,\\
\breve{\varepsilon}^{13} &= \varepsilon^{13}(\varepsilon^{33})^{-1}\,,\\
\breve{\varepsilon}^{21} &= \varepsilon^{21} - \varepsilon^{23}(\varepsilon^{33})^{-1}\varepsilon^{31}\,,\\
\breve{\varepsilon}^{22} &= \varepsilon^{22} - \varepsilon^{23}(\varepsilon^{33})^{-1}\varepsilon^{32}\,,\\
\breve{\varepsilon}^{23} &= \varepsilon^{23}(\varepsilon^{33})^{-1}\,,\\
\breve{\varepsilon}^{31} &= (\varepsilon^{33})^{-1}\varepsilon^{31}\,,\\
\breve{\varepsilon}^{32} &= (\varepsilon^{33})^{-1}\varepsilon^{32}\,,\\
\breve{\varepsilon}^{33} &= (\varepsilon^{33})^{-1}\,.
\end{aligned} \tag{5.6}$$

Similar expressions for the permeability $\underline{\breve{\boldsymbol{\mu}}}$ are obtained by replacing $\varepsilon \to \mu$.

We solve the remaining four equations in Eqs. (5.4) for the terms containing derivatives ∂_3, and eliminate all occurrences of the third field components with help of Eqs. (5.5). Hence, we get a set of four coupled equations containing the transversal field components only. Introducing the electromagnetic field component vector $\mathbf{V} = (-E_2, E_1, H_1, H_2)^T$, we can rewrite Maxwell's equations into a first order differential matrix-vector equation

$$\frac{\partial_3}{\mathrm{i}}\begin{pmatrix} -E_2 \\ E_1 \\ H_1 \\ H_2 \end{pmatrix} = \underline{\boldsymbol{\mathcal{A}}}\begin{pmatrix} -E_2 \\ E_1 \\ H_1 \\ H_2 \end{pmatrix} \tag{5.7}$$

where the system matrix operator $\underline{\boldsymbol{\mathcal{A}}}$ contains the derivatives in transversal directions. Explicitly, the

[3] Please note that we use the negative $\left(-E_2\right)$ component instead of E_2 in the following, because this is the way our numerical framework has historically been implemented, even though there are no obvious reasons for this convention. Still, we stick to this notational artifact in order to be comparable to the code.

system matrix operator is given by

$$\underline{\boldsymbol{\mathcal{A}}} = \begin{pmatrix} -\frac{\partial_2}{\mathrm{i}}\check{\varepsilon}^{32} - \check{\mu}^{13}\frac{\partial_1}{\mathrm{i}} & \frac{\partial_2}{\mathrm{i}}\check{\varepsilon}^{31} - \check{\mu}^{13}\frac{\partial_2}{\mathrm{i}} & \omega^2\check{\mu}^{11} - \frac{\partial_2}{\mathrm{i}}\check{\varepsilon}^{33}\frac{\partial_2}{\mathrm{i}} & \omega^2\check{\mu}^{12} + \frac{\partial_2}{\mathrm{i}}\check{\varepsilon}^{33}\frac{\partial_1}{\mathrm{i}} \\ \frac{\partial_1}{\mathrm{i}}\check{\varepsilon}^{32} - \check{\mu}^{23}\frac{\partial_1}{\mathrm{i}} & -\frac{\partial_1}{\mathrm{i}}\check{\varepsilon}^{31} - \check{\mu}^{23}\frac{\partial_2}{\mathrm{i}} & \omega^2\check{\mu}^{21} + \frac{\partial_1}{\mathrm{i}}\check{\varepsilon}^{33}\frac{\partial_2}{\mathrm{i}} & \omega^2\check{\mu}^{22} - \frac{\partial_1}{\mathrm{i}}\check{\varepsilon}^{33}\frac{\partial_1}{\mathrm{i}} \\ \check{\varepsilon}^{22} - \frac{1}{\omega^2}\frac{\partial_1}{\mathrm{i}}\check{\mu}^{33}\frac{\partial_1}{\mathrm{i}} & -\check{\varepsilon}^{21} - \frac{1}{\omega^2}\frac{\partial_1}{\mathrm{i}}\check{\mu}^{33}\frac{\partial_2}{\mathrm{i}} & -\check{\varepsilon}^{23}\frac{\partial_2}{\mathrm{i}} - \frac{\partial_1}{\mathrm{i}}\check{\mu}^{31} & \check{\varepsilon}^{23}\frac{\partial_1}{\mathrm{i}} - \frac{\partial_1}{\mathrm{i}}\check{\mu}^{32} \\ -\check{\varepsilon}^{12} - \frac{1}{\omega^2}\frac{\partial_2}{\mathrm{i}}\check{\mu}^{33}\frac{\partial_1}{\mathrm{i}} & \check{\varepsilon}^{11} - \frac{1}{\omega^2}\frac{\partial_2}{\mathrm{i}}\check{\mu}^{33}\frac{\partial_2}{\mathrm{i}} & \check{\varepsilon}^{13}\frac{\partial_2}{\mathrm{i}} - \frac{\partial_2}{\mathrm{i}}\check{\mu}^{31} & -\check{\varepsilon}^{13}\frac{\partial_1}{\mathrm{i}} - \frac{\partial_2}{\mathrm{i}}\check{\mu}^{32} \end{pmatrix}, \tag{5.8}$$

and can be decomposed into four 2x2 submatrices as follows:

$$\underline{\boldsymbol{\mathcal{A}}} = \begin{pmatrix} \underline{\boldsymbol{\mathcal{A}}}_{11} & \underline{\boldsymbol{\mathcal{F}}} \\ \underline{\boldsymbol{\mathcal{G}}} & \underline{\boldsymbol{\mathcal{A}}}_{22} \end{pmatrix}. \tag{5.9}$$

The reason for highlighting the two submatrices denoted with $\underline{\boldsymbol{\mathcal{F}}}$ and $\underline{\boldsymbol{\mathcal{G}}}$ will become apparent in a moment.

Large Eigenvalue Problem

By virtue of the staircase approximation each layer is homogeneous along the primary x^3-direction: within the layer, $\underline{\varepsilon}$ and $\underline{\mu}$ only depend on x^1 and x^2 coordinates and not on x^3. This allows for a plane wave ansatz $e^{\mathrm{i}\gamma x^3}$ for the fields, where $\gamma = k_3 = k_z$ is the propagation constant, the third component of the wave vector along the primary direction. Consequently, the derivative ∂_3 can be replaced by $\mathrm{i}\gamma$, and we obtain the *large eigenvalue equation*

$$\gamma \begin{pmatrix} -E_2 \\ E_1 \\ H_1 \\ H_2 \end{pmatrix} = \underline{\boldsymbol{\mathcal{A}}} \begin{pmatrix} -E_2 \\ E_1 \\ H_1 \\ H_2 \end{pmatrix}. \tag{5.10}$$

It contains all transversal field components of the eigenmode as eigenvectors and the corresponding propagation constant as the related eigenvalue. Once an eigenvector is known, the longitudinal field components can be calculated from Eqs. (5.5). There is an infinite number of solutions to this equation. As long as we have reciprocal materials, i.e., permittivity and permeability tensors are symmetric, the j-th eigensolution with eigenvalue γ_j is accompanied by an equivalent eigensolution k with eigenvalue $\gamma_k = -\gamma_j$. Hence, there always exist two associated electromagnetic waves — one mode traveling in the positive x^3-direction (forward) and the same mode traveling in the negative x^3-direction (backward).

Small Eigenvalue Problem

The coupled equations of the large eigenvalue problem can be partially decoupled if the full anisotropy of permittivity and permeability tensors is not needed. Examining the system matrix operator in Eq. (5.8) and Eq. (5.9), it becomes apparent that $\underline{\boldsymbol{\mathcal{A}}}_{11}$ and $\underline{\boldsymbol{\mathcal{A}}}_{22}$ vanish if the permittivity is

restricted to anisotropy in the x^1-x^2-plane

$$\underline{\boldsymbol{\varepsilon}} = \begin{pmatrix} \varepsilon^{11} & \varepsilon^{12} & 0 \\ \varepsilon^{21} & \varepsilon^{22} & 0 \\ 0 & 0 & \varepsilon^{33} \end{pmatrix} \quad \xrightarrow{\text{Eq. (5.6)}} \quad \underline{\check{\boldsymbol{\varepsilon}}} = \begin{pmatrix} \varepsilon^{11} & \varepsilon^{12} & 0 \\ \varepsilon^{21} & \varepsilon^{22} & 0 \\ 0 & 0 & (\varepsilon^{33})^{-1} \end{pmatrix} . \tag{5.11}$$

For the permittivity $\underline{\boldsymbol{\mu}}$ and $\underline{\check{\boldsymbol{\mu}}}$ similar restrictions apply.

This form of anisotropy is important since it allows for in-plane coordinate transformations used by adaptive meshing techniques and stretched coordinate perfectly matched layers, which we are going to introduce in Chap. 7. However, the submatrices $\underline{\boldsymbol{\mathcal{F}}}$ and $\underline{\boldsymbol{\mathcal{G}}}$ remain as before, and Eq. (5.10) can be separated into two sets of coupled equations

$$\frac{\partial_3}{\mathrm{i}} \begin{pmatrix} -E_2 \\ E_1 \end{pmatrix} = \underline{\boldsymbol{\mathcal{F}}} \begin{pmatrix} H_1 \\ H_2 \end{pmatrix} , \tag{5.12a}$$

$$\frac{\partial_3}{\mathrm{i}} \begin{pmatrix} H_1 \\ H_2 \end{pmatrix} = \underline{\boldsymbol{\mathcal{G}}} \begin{pmatrix} -E_2 \\ E_1 \end{pmatrix} . \tag{5.12b}$$

By multiplying one of these equations with $\frac{\partial_3}{\mathrm{i}}$ and substituting the other — and vice versa — the electric and magnetic fields can be decoupled. Then, replacing $\frac{\partial_3}{\mathrm{i}} \to \gamma$ by virtue of the plane wave ansatz, both variants read

$$\gamma^2 \begin{pmatrix} -E_2 \\ E_1 \end{pmatrix} = \underline{\boldsymbol{\mathcal{F}}}\,\underline{\boldsymbol{\mathcal{G}}} \begin{pmatrix} -E_2 \\ E_1 \end{pmatrix} , \tag{5.13a}$$

$$\gamma^2 \begin{pmatrix} H_1 \\ H_2 \end{pmatrix} = \underline{\boldsymbol{\mathcal{G}}}\,\underline{\boldsymbol{\mathcal{F}}} \begin{pmatrix} H_1 \\ H_2 \end{pmatrix} . \tag{5.13b}$$

These eigenvalue equations are half of the size of the large eigenvalue equation and formulated for either the transversal electric or transversal magnetic field components. Consequently, these eigenvalue equations are called *small eigenvalue equations*. Since the eigenmodes are fully described by either the electric or the magnetic fields, only of of these equations needs to be solved. If we choose to solve for the E-fields with Eq. (5.13a), the H-fields are obtained from Eq. (5.12b) divided by the propagation constant γ. Accordingly, if we choose to solve for the H-fields with Eq. (5.13b), the E-fields are obtained from Eq. (5.12a) divided by the propagation constant.

The eigenvalues in Eqs. (5.13) are now the squared propagation constants. Hence, the propagation constants are obtained by taking the square roots of the eigenvalues. In contrast to the large eigenvalue problem, each eigenmode appears only once. However, similar to before, the electromagnetic waves appear as equivalent forward and backward traveling modes with the positive and negative solution of the eigenvalue's (complex) square root $\pm\sqrt{\gamma^2}$, respectively.

The advantage of the small eigenvalue problem is that it is half the size of the large problem. This reduction by a factor of two speeds up the numerical solution of the problem roughly by a factor of eight, because the computational time for the diagonalization of a general dense matrix operator

is in the order of $\mathcal{O}(N^3)$ where N denotes the dimension of the matrix operator. Furthermore, the obligatory sorting of the eigenmodes into forward and backward traveling modes is not necessary.

Further special cases like diagonal or isotropic material tensors can easily be deduced by setting those tensor entries to zero which are not needed.

Numerical Discretization of the Eigenmodes

In the last sections we derived large and small eigenvalue equations for the eigenmodes of the l-th layer. Each layer has infinitely many eigenmodes. When we calculate them numerically, however, we have to discretize the eigenvalue equations. For the numerical representation of each field component we use a set of M basis functions $\{\mathcal{B}_m, m = 1, \ldots, M\}$, e.g., for the numerical representation E_1 of the first electric field component E_1 we have

$$E_1(x^1, x^2) \approx \mathsf{E}_1(x^1, x^2) = \sum_{m=1}^{M} \tilde{\mathsf{E}}_{1,m}\, \mathcal{B}_m(x^1, x^2)\,., \tag{5.14}$$

The details of such an expansion with a concrete set of basis functions and the corresponding discretization of the matrix operator $\underline{\mathbf{A}}$ will be topic of Chap. 6. Here, we focus on the principle that, independently of the choice of a particular basis, the field components can be represented by a vector of associated expansion coefficients

$$\tilde{\mathbf{E}}_1 = (\tilde{\mathsf{E}}_{1,1}, \ldots, \tilde{\mathsf{E}}_{1,m}, \ldots, \tilde{\mathsf{E}}_{1,M})^T\,, \tag{5.15}$$

where the used sans serif symbols denote the numerical representation character and the tilde symbol $\tilde{\cdot}$ the discretized version of the quantity, i.e., the expansion coefficients or the coefficient vector. The same discretization applies to all other field components. Hence, their numerical representation is given by $-\mathsf{E}_2(x^1, x^2)$, $\mathsf{H}_1(x^1, x^2)$, and $\mathsf{H}_2(x^1, x^2)$, or alternatively the coefficient vectors $(-\tilde{\mathbf{E}}_2)$, $\tilde{\mathbf{H}}_1$, and $\tilde{\mathbf{H}}_2$. The coefficient vector notation will become more important later. For now, it is sufficient to realize that the discretization leaves $2M$ degrees of freedom for the small eigenproblem and $4M$ degrees of freedom for the large eigenproblem. As a direct consequence, the discretized eigenproblems provide only finite subsets of the continuous problem's infinite set of eigenmodes. These finite subsets comprises $N_j^s = 2M = \bar{N}$ and $N_j^l = 4M = 2\bar{N}$ independent eigenmodes for the small eigenvalue equations, Eqs. (5.13), and large eigenvalue equation, Eq. (5.10), respectively. However, in the case of the small eigenproblem, we also get a total of $4M$ solutions due to the positive and negative roots of the eigenvalues $\pm\sqrt{\gamma^2}$ corresponding to forward and backward traveling eigenmodes. Note, that the numerical eigenmodes fulfill the original eigenvalue equations only approximately, i.e., Eq. (5.10) is replaced by

$$\underline{\boldsymbol{\mathcal{A}}} \begin{pmatrix} -\mathsf{E}_2 \\ \mathsf{E}_1 \\ \mathsf{H}_1 \\ \mathsf{H}_2 \end{pmatrix}_j \approx \gamma_j \begin{pmatrix} -\mathsf{E}_2 \\ \mathsf{E}_1 \\ \mathsf{H}_1 \\ \mathsf{H}_2 \end{pmatrix}_j\,, \tag{5.16}$$

with subscript labels $j = 1, \ldots, N_j^l$ denoting the different eigenmodes.[4] Only in the limit $M \to \infty$ the numerical eigenmodes are exact.

[4] The small eigenproblems, $Eqs.$ (5.13), can be rewritten in a similar way with mode labels $j = 1 \ldots, N_j^s$.

5.2.2. Field Expansion

We use the calculated set of eigenmodes of a layer as an expansion basis for the fields in this layer. Since the descriptions for the small and large eigenproblems are slightly different, we will separate their presentations.

Expansion in the Small Eigenproblem Case

By definition, the small eigenproblem always provides as many forward traveling modes as backward traveling modes. The associated forward and backward modes have the same label j but different signs in front of the eigenvalue γ_j. In the exemplary case of eigenproblem Eq. (5.13a), the transversal electric fields $\mathbf{E}_\parallel = (-E_2, E_1)^T$ in the l-layer are expanded into the obtained eigenmodes as

$$\mathbf{E}_\parallel(x^1, x^2, x^3) = \sum_{j=1}^{\bar{N}} \Big[\underbrace{u_j \, e^{\mathrm{i}\gamma_j (x^3 - x^3_{l-1})}}_{\substack{\text{forward} \\ \text{traveling}}} + \underbrace{d_j \, e^{-\mathrm{i}\gamma_j (x^3 - x^3_l)}}_{\substack{\text{backward} \\ \text{traveling}}} \Big] \underbrace{\mathbf{E}_{\parallel j}(x^1, x^2)}_{\substack{\text{calculated} \\ \text{eigenmode}}} \tag{5.17}$$

for $x^3_{l-1} \leq x^3 \leq x^3_l$, where we have used the electric field eigenmode vector $\mathbf{E}_{\parallel j} = (-\mathsf{E}_2, \mathsf{E}_1)^T_j$ of the j-th eigenmode. The new expansion coefficients u_j and d_j are still unknown and will be calculated using the scattering matrix algorithm derived in Sec. 5.3. Of course, the eigenmode expansion incorporates the plane wave ansatz for the x^3-dependence. The phase offset x^3_{l-1} in forward plane waves provides for the layer's displacement with respect to the origin and gives the coordinate relative to the layer's backward boundary. Similarly, the offset x^3_l in the backward phase ensures that backward traveling modes start at the forward boundary (cf. Fig. 5.2) [24].

Furthermore, as long as the imaginary part of the propagation constant $\mathrm{Im}(\gamma_j) \geq 0$, the backward eigenmodes actually travel backward and numerical stability is improved because the exponential terms do not grow and small numerical errors are never amplified. We will build upon this in the scattering matrix algorithm. These considerations fix a rule which square root to pick for γ_j: We pick the one with $\mathrm{Im}(\gamma_j) > 0$, or with $\mathrm{Re}(\gamma_j) > 0$ if the imaginary part vanishes.[5] Then the forward traveling modes with amplitude u_j are either forward propagating, forward damped, or forward evanescent, and the backward traveling modes with amplitude d_j are either backward propagating, backward damped, or backward evanescent.

The magnetic field is obtained by substituting the transversal electric field components of Eq. (5.17) into Eq. (5.12b). The transversal magnetic fields $\mathbf{H}_\parallel = (H_1, H_2)^T$ then similarly read

$$\mathbf{H}_\parallel(x^1, x^2, x^3) = \sum_{j=1}^{\bar{N}} \Big[\underbrace{+ u_j \, e^{\mathrm{i}\gamma_j (x^3 - x^3_{l-1})}}_{=: u_j(x^3)} \underbrace{- d_j \, e^{-\mathrm{i}\gamma_j (x^3 - x^3_l)}}_{=: d_j(x^3)} \Big] \mathbf{H}_{\parallel j}(x^1, x^2) \tag{5.18}$$

for $x^3_{l-1} \leq x^3 \leq x^3_l$, where we have used the magnetic field eigenmode vector $\mathbf{H}_{\parallel j} = (\mathsf{H}_1, \mathsf{H}_2)^T_j$ of the j-th eigenmode. The latter is obtained from the numerical analog of Eq. (5.12b),

$$\mathbf{H}_{\parallel j} = \frac{1}{\gamma_j} \underline{\boldsymbol{\mathcal{G}}} \, \mathbf{E}_{\parallel j} \, . \tag{5.19}$$

[5] Li [51] describes a different rule to pick the square root of $\pm\sqrt{\gamma_j^2}$. He allows for small negative imaginary parts by using the rule $\mathrm{Re}(\gamma_j) + \mathrm{Im}(\gamma_j) > 0$.

The red positive and negative signs in Eq. (5.18) are provided by the ∂_3 derivative of forward and backward traveling modes, respectively. The missing $E_3(x^1, x^2, x^3)$ and $H_3(x^1, x^2, x^3)$ components are obtained from the substitution of $\mathbf{E}_\parallel$ and $\mathbf{H}_\parallel$ into Eqs. (5.5).[6]

Expansion in the Large Eigenproblem Case

Large eigenmodes can be used as an expansion basis as well. In contrast to the small eigenproblem, the large eigenproblem does not guarantee that we always get a forward mode and its corresponding backward mode. Even though it is usually not the case, it may happen that, due to the numerical treatment, only one of the equivalent modes is calculated. To avoid numerical instability, we would like to distinguish between forward and backward modes and treat them accordingly as in the small eigenproblem. As it makes life much easier later on, we use as many forward as backward traveling modes. Hence, it it necessary to sort the modes into two groups. We use the scheme noted in Tab. 5.1 to obtain the forward modes with propagation constant γ_j^+ and the backward modes with eigenvalues γ_j^-. The fields $\mathbf{V} = (\mathbf{E}_\parallel, \mathbf{H}_\parallel)^T$ in the l-th layer can then be written as an expansion into forward ($\mathbf{V}_j^+$) and backward ($\mathbf{V}_j^-$) traveling numerical eigenmodes

$$\begin{aligned}\mathbf{V}(x^1, x^2, x^3) &\approx \mathsf{V}(x^1, x^2, x^3) \\ &= \sum_{j'=1}^{2\bar{N}} a_{j'}(x^3)\, \mathsf{V}_{j'}(x^1, x^2) \\ &= \sum_{j=1}^{\bar{N}} \underbrace{u_j\, e^{\mathrm{i}\gamma_j^+ (x^3 - x_{l-1}^3)}\, \mathsf{V}_j^+(x^1, x^2)}_{\substack{\text{forward traveling} \\ \text{eigenmodes}}} + \underbrace{d_j\, e^{\mathrm{i}\gamma_j^- (x^3 - x_l^3)}\, \mathsf{V}_j^-(x^1, x^2)}_{\substack{\text{backward traveling} \\ \text{eigenmodes}}}\,, \end{aligned} \tag{5.20}$$

for $x_{l-1}^3 \leq x^3 \leq x_l^3$, with $\mathsf{V}_j = (\mathsf{E}_\parallel, \mathsf{H}_\parallel)_j^T$. The expansion amplitudes u_j and d_j are still unknown and need to be determined with the scattering matrix algorithm presented in Sec. 5.3. We can al-

[6] An alternative way is the calculation of E_3 and H_3 components from $\mathsf{E}_\parallel$ and $\mathsf{H}_\parallel$ directly. Then the entire field vectors can be expanded similarly to Eq. (5.17) and Eq. (5.18).

$\mathrm{Re}(\gamma_j)$	$\mathrm{Im}(\gamma_j)$	Group	
> 0	$= 0$	forward propagating	γ_j^+
≥ 0	> 0	forward decaying	
< 0	> 0	forward mixed	
< 0	$= 0$	backward propagating	γ_j^-
≤ 0	< 0	backward decaying	
> 0	< 0	backward mixed	
$= 0$	$= 0$	distribute to even out size of groups	

Table 5.1.: *Sorting scheme for the eigenmodes of the large eigenvalue problem.*

ways distinguish between backward and forward modes in the expansion as above. However, the particular notation of the third line in Eq. (5.20), where forward and backward modes have the same label j, is only advisable in the case where modes with the same label are the corresponding forward and backward traveling representations of the *same* mode. As stated before, in contrast to the small eigenproblem, this is not guaranteed in the large eigenproblem. Instead, even the number of forward and backward modes could be different — only the total number of eigensolutions is fixed. Nevertheless, we use this notation for the convenience of a simple illustration.

Matrix Vector Notation

The field expansions Eq. (5.17), Eq. (5.18), and Eq. (5.20) provide the fields in the layer l if we know all expansion coefficients $u_j^{(l)}$ and $d_j^{(l)}$. We would like to separate these unknowns from the known rest of the equations, i.e., from the eigenmodes and the phase factors. This is conveniently achieved by writing all expansion amplitudes of the l-th layer in vectors[7]

$$\mathbf{u}^{(l)} = (u_1^{(l)}, \ldots, u_j^{(l)}, \ldots, u_{\bar{N}}^{(l)})^T \,, \tag{5.21a}$$

$$\mathbf{d}^{(l)} = (d_1^{(l)}, \ldots, d_j^{(l)}, \ldots, d_{\bar{N}}^{(l)})^T \,. \tag{5.21b}$$

The field vector $\mathbf{V}^{(l)}$ (which comprises all transverse electric and magnetic field components) can then be written as a product of a matrix $\underline{\mathbf{M}}^{(l)}$ containing the eigenmodes (in columns), a phase matrix $\underline{\mathbf{\Phi}}^{(l)}$ containing the exponential terms, and the new amplitude vectors:

$$\mathbf{V}^{(l)}(x^1,x^2,x^3) = \underbrace{\underline{\mathbf{M}}^{(l)}(x^1,x^2)}_{\substack{\text{eigenmode}\\ \text{matrix}}} \underbrace{\underline{\mathbf{\Phi}}^{(l)}(x^3)}_{\substack{\text{phase}\\ \text{matrix}}} \underbrace{\begin{pmatrix}\mathbf{u}^{(l)}\\ \mathbf{d}^{(l)}\end{pmatrix}}_{\substack{\text{amplitude}\\ \text{vector}}} = \underline{\mathbf{M}}^{(l)}(x^1,x^2) \underbrace{\begin{pmatrix}\mathbf{u}^{(l)}(x^3)\\ \mathbf{d}^{(l)}(x^3)\end{pmatrix}}_{\substack{\text{phased amplitude}\\ \text{vector}}} . \tag{5.22}$$

In some situations it is convenient to combine phase matrix and amplitude vector to form the phased amplitude vector at a certain coordinate x^3.

The eigenmode matrix for small and large eigenproblems look slightly different. The former is given by

$$\underline{\mathbf{M}}^{(l)} = \begin{pmatrix} \underline{\mathbf{M}}_{\mathsf{E}}^{(l)} & \underline{\mathbf{M}}_{\mathsf{E}}^{(l)} \\ \underline{\mathbf{M}}_{\mathsf{H}}^{(l)} & -\underline{\mathbf{M}}_{\mathsf{H}}^{(l)} \end{pmatrix} , \tag{5.23}$$

where the $2 \times \bar{N}$ submatrices consist of all eigenmode column vectors written in a row

$$\underline{\mathbf{M}}_{\mathsf{E}}^{(l)} = \left(\mathbf{E}_{\parallel 1}^{(l)}, \mathbf{E}_{\parallel 2}^{(l)}, \ldots \right) , \qquad \underline{\mathbf{M}}_{\mathsf{H}}^{(l)} = \left(\mathbf{H}_{\parallel 1}^{(l)}, \mathbf{H}_{\parallel 2}^{(l)}, \ldots \right) . \tag{5.24}$$

The eigenmode matrix for the large eigenproblem appears as

$$\underline{\mathbf{M}}^{(l)} = \left(\underline{\mathbf{V}}^{(l)+} \;\; \underline{\mathbf{V}}^{(l)-} \right) , \tag{5.25}$$

[7] The general case is described by $\mathbf{u}^{(l)} = (a_1^{(l)}, \ldots, a_{j'}^{(l)}, \ldots, a_{\bar{N}^+}^{(l)})^T$ and $\mathbf{d}^{(l)} = (a_{\bar{N}^++1}^{(l)}, \ldots, a_{j'}^{(l)}, \ldots, a_{2\bar{N}}^{(l)})^T$, where we assume that the eigenmodes are sorted such that the first $\bar{N}^+$ modes with $1 \leq \bar{N}^+ < 2\bar{N}$ are forward modes and the rest are backward modes. However, the number of forward traveling modes $\bar{N}^+ \approx \bar{N}$ is usually in the order of half the total number of modes.

where the $4 \times \bar{N}$ submatrices consist of the forward or backward eigenmode column vectors written in a row

$$\underline{\mathbf{V}}^{(l)\pm} = \left(\mathbf{V}_1^{(l)\pm}, \mathbf{V}_2^{(l)\pm}, \ldots, \mathbf{V}_{\bar{N}}^{(l)\pm} \right) . \tag{5.26}$$

The phase matrix $\underline{\mathbf{\Phi}}^{(l)}(x^3)$ is a diagonal matrix with entries

$$\underline{\mathbf{\Phi}}^{(l)}(x^3) = \begin{pmatrix} \underline{\mathbf{\Phi}}^{(l)+} & 0 \\ 0 & \underline{\mathbf{\Phi}}^{(l)-} \end{pmatrix} \tag{5.27}$$

where the diagonal submatrices can be written as

$$\underline{\mathbf{\Phi}}^{(l)+}(x^3) = \operatorname{diag}\left(e^{\mathrm{i}\gamma_1^+(x^3 - x_{l-1}^3)}, \ldots, e^{+\mathrm{i}\gamma_{\bar{N}}^+(x^3 - x_{l-1}^3)} \right) , \tag{5.28a}$$

$$\underline{\mathbf{\Phi}}^{(l)-}(x^3) = \operatorname{diag}\left(e^{\mathrm{i}\gamma_1^-(x^3 - x_l^3)}, \ldots, e^{\mathrm{i}\gamma_{\bar{N}}^-(x^3 - x_l^3)} \right) \tag{5.28b}$$

for the large eigenproblem[8], and similarly with $\gamma_j^+ = \gamma_j$ and $\gamma_j^- = -\gamma_j$ for the small eigenproblem.

Up to here, we have decomposed the structure into layers homogeneous in x^3-direction. This staircase approximation allowed us to calculate a set of eigenmodes in each layer which we used as an expansion basis for the layer's fields. The field expansion could be rewritten into a matrix notation. The remaining unknowns are the expansion amplitude vectors $\mathbf{u}^{(l)}$ and $\mathbf{d}^{(l)}$ in every single layer. Hence, the next step is to reduce the $L \cdot 2\bar{N}$ unknown amplitude coefficients to $2 \cdot 2\bar{N}$ unknowns – $2\bar{N}$ in the first and last layer each – by reconnecting the layers via the electromagnetic field continuity conditions. This is the purpose of the scattering matrix algorithm.

5.3. Scattering Matrix Algorithm

The *scattering matrix* (S-matrix) in general describes the scattering solutions of a system. It is an operator which connects the initial state to the final state of a physical system given as superpositions of scattering channels. The elements of the matrix operator — known as scattering amplitudes — can be interpreted as transition probabilities between the initial and final channels.

We use the scattering matrix to describe repeated multiple diffraction of electromagnetic waves at the interfaces between adjacent layers of a staircase approximated structure. The scattering channels are the eigenmodes of the respective layers represented by the expansion amplitude vectors $\mathbf{u}^{(l)}$ and $\mathbf{d}^{(l)}$, and the scattering strength (transition probability) is calculated from the field continuity conditions and the corresponding field matching at the interfaces. The field matching can be done either pointwise, in form of an overlap integral of the eigenmodes, or by matching the basis functions directly. All three cases are discussed in Sec. 5.3.1.

In contrast to the common *transfer matrix* (T-matrix) approach, which is often used to propagate fields through a layered structure, the S-matrix does not propagate the fields from left to right as schematically depicted for the T-matrix in Fig. 5.3(b). The defining equation for the T-matrix is

[8] The numbering of the phase matrix entries corresponding to the large eigensolutions is in principle the same as that of the amplitudes (cf. footnote 7). This means,
$\underline{\mathbf{\Phi}}^{(l)+} = \operatorname{diag}[e^{\mathrm{i}\gamma_1(x^3 - x_{l-1}^3)}, \ldots, e^{+\mathrm{i}\gamma_{\bar{N}^+}(x^3 - x_{l-1}^3)}]$ and $\underline{\mathbf{\Phi}}^{(l)-} = \operatorname{diag}[e^{\mathrm{i}\gamma_{\bar{N}+1}(x^3 - x_l^3)}, \ldots, e^{+\mathrm{i}\gamma_{2\bar{N}}(x^3 - x_l^3)}]$.

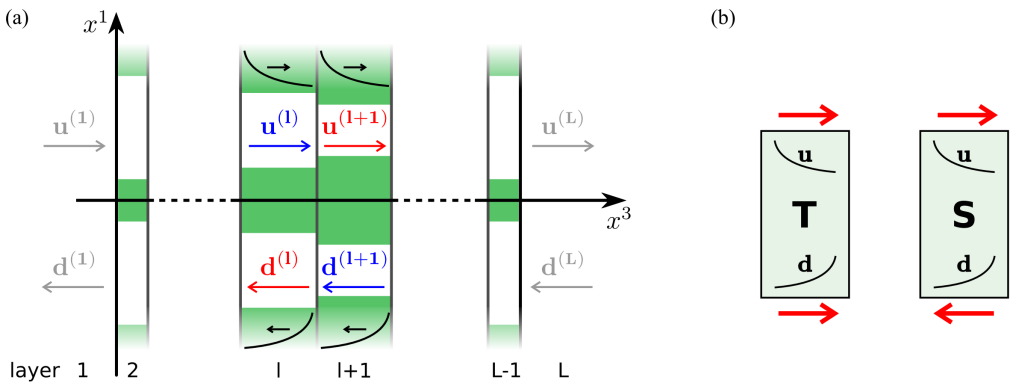


Figure 5.3.: *(a) Illustration of forward and backward amplitudes in the system. The incoming (blue) and outgoing (red)amplitudes are depicted with respect to the interface between layers l and $l+1$. The small black curves and arrows above and below indicate the direction of their decay. (b) Schematic comparison between T- and S-matrix approaches. T-matrices propagate both* $\mathbf{u}$ *and* $\mathbf{d}$ *amplitudes from left to right — the latter against their direction of decay. S-matrices propagate* $\mathbf{u}$ *and* $\mathbf{d}$ *amplitudes in opposite directions — along their directions of decay.*

$$\begin{pmatrix} \mathbf{u}^{(l+1)} \\ \mathbf{d}^{(l+1)} \end{pmatrix} = \underline{\mathbf{T}}(l, l+1) \begin{pmatrix} \mathbf{u}^{(l)} \\ \mathbf{d}^{(l)} \end{pmatrix} , \tag{5.29}$$

where we observe that incoming amplitudes and outgoing amplitudes appear together in the vectors. Thus, backward modes must be propagated *against* their direction of potentially decaying phase factors.

In contrast to this, the S-matrix propagates the eigenmodes along their natural traveling direction — forward modes forwards, backward modes backwards (cf. Fig. 5.3 (b)) — particularly the evanescent modes. As a direct consequence, the evanescent modes are propagated along the direction of their natural decay such that the amplitudes always diminish. Thus, numerical instability due to *exponentially growing* phase factors is avoided.[9] The topic of stability is extensively discussed in Ref. [76]. The S-matrix $\underline{\mathbf{S}}(l, l+1)$ connecting the incoming amplitudes to the outgoing amplitudes in layers l and $l+1$ is given by

$$\begin{pmatrix} \mathbf{u}^{(l+1)} \\ \mathbf{d}^{(l)} \end{pmatrix} = \underline{\mathbf{S}}(l, l+1) \begin{pmatrix} \mathbf{u}^{(l)} \\ \mathbf{d}^{(l+1)} \end{pmatrix} . \tag{5.30}$$

The downside of the S-matrix stability is a complicated procedure to obtain compositions of S-matrices. The composition (product) of T-matrices is given by the ordinary matrix product

$$\underline{\mathbf{T}}(l, l+2) = \underline{\mathbf{T}}(l+1, l+2) \cdot \underline{\mathbf{T}}(l, l+1) , \tag{5.31}$$

whereas the scattering matrix product is given by

$$\underline{\mathbf{S}}(l, l+2) = \underline{\mathbf{S}}(l+1, l+2) \star \underline{\mathbf{S}}(l, l+1) . \tag{5.32}$$

[9] These instabilities appear as exploding transmittance and reflectance coefficients and, thus, violation of energy conservation, when the differences in the expansion amplitudes' magnitudes exceeds the numerical accuracy.

This S-matrix product, denoted by the $\star$ symbol, is cumbersome and will be introduced in Sec. 5.3.3.

In general, scattering matrices can be decomposed into four submatrices

$$\underline{\mathbf{S}}(l, l+1) = \begin{pmatrix} \underline{\mathbf{S}}_{uu} & \underline{\mathbf{S}}_{ud} \\ \underline{\mathbf{S}}_{du} & \underline{\mathbf{S}}_{dd} \end{pmatrix} . \tag{5.33}$$

These submatrices have a physical interpretation. For example, $\underline{\mathbf{S}}_{uu}$ transforms $\mathbf{u}^{(l)}$ into $\mathbf{u}^{(l+1)}$ (cf. Eq. (5.30)). Thus, it describes the propagation of the forward eigenmodes of layer l through this layer and the coupling across the interface between layers l and $l+1$ to the eigenmodes of layer $l+1$ — in short: the transmission from layer l to $l+1$. Likewise, submatrix $\underline{\mathbf{S}}_{ud}$ connects amplitudes $\mathbf{d}^{(l+1)}$ with $\mathbf{u}^{(l+1)}$ which describes the reflectance at the same interface from the backward eigenmodes into the forward eigenmodes within layer $l+1$. The other two submatrices can be interpreted accordingly. If an S-matrix describes a larger part of the structure, i.e., a set of adjacent layers l through $l+p$, $p > 1$, the interpretation remains the same. However, the scattering matrix $\underline{\mathbf{S}}(l, l+p)$ additionally (automatically) includes the entire multitude of multiple reflection and transmission (diffraction) processes at all intermediate interfaces. This means, the scattering matrix includes all possible paths the light can take.

In order to make our lives easier, we restrict the number of modes we use in the scattering matrix algorithm in every layer to $\bar{N}$ forward and $\bar{N}$ backward modes. Then the submatrices of the S-matrix are quadratic $\bar{N} \times \bar{N}$ matrices and the scattering matrix contains $2\bar{N} \times 2\bar{N}$ entries.

After having introduced the general properties of scattering matrices, the next topic is the derivation of the expressions necessary to calculate the scattering matrix connecting adjacent layers. The S-matrix algorithm describes how to obtain the scattering matrix $\underline{\mathbf{S}}(l, l+p)$ from the eigensolutions of the individual layers involved. It builds upon the basic procedure to transform an amplitude vector, e.g., $\mathbf{u}^{(l)}$, into the amplitude vector $\mathbf{u}^{(l+1)}$ in the following layer. This process comprises two steps: First, a propagation through the layer l, and, second, the crossing of the interface between layers l and $l+1$. Repetition of this sequence in the proper fashion for every additional layer and amplitude gives the desired S-matrix.

The *total scattering matrix* describes the scattering response of the whole structure and connects the amplitudes of the first layer 1 with those of the last layer L like

$$\begin{pmatrix} \mathbf{u}^{(L)} \\ \mathbf{d}^{(1)} \end{pmatrix} = \underline{\mathbf{S}}(1, L) \begin{pmatrix} \mathbf{u}^{(1)} \\ \mathbf{d}^{(L)} \end{pmatrix} . \tag{5.34}$$

If we specify the amplitudes of the incident eigenmodes by an expansion of the incoming waves on both sides, we can calculate the outgoing waves from $\mathbf{u}^{(L)}$ and $\mathbf{d}^{(1)}$. In particular, we usually have only incidence from one side, e.g., $\mathbf{u}^{(1)} \neq \mathbf{0}$ and $\mathbf{d}^{(L)} = \mathbf{0}$. Then, we obtain the amplitudes of the transmitted and reflected fields from

$$\text{transmitted amplitudes:} \qquad \mathbf{u}^{(L)} = \underline{\mathbf{S}}_{uu}(1, L)\, \mathbf{u}^{(1)} , \tag{5.35a}$$

$$\text{reflected amplitudes:} \qquad \mathbf{d}^{(1)} = \underline{\mathbf{S}}_{du}(1, L)\, \mathbf{u}^{(1)} . \tag{5.35b}$$

Before we advance to the detailed description of the scattering matrix algorithm in Sec. 5.3.2, we have a look at how the diffraction into the different eigenmodes actually incorporates through the compliance of the field continuity conditions at interfaces of adjacent layers. The interface matrix, which we will derive next, is the essential ingredient to the scattering matrix recursion.

5.3.1. Field Matching at Interfaces

The matching of the continuous electromagnetic fields at the interface of adjacent layers is the key step to recombine the L layer subsystems to the full three-dimensional structure via the scattering matrix algorithm. As stated above, this field matching reduces the degrees of freedom to the number that can be provided by the incident waves expanded into the first and last layer's eigenmodes.

In this section, we show how the matching is achieved. To this end, we consider the interface at x_l^3 between layers l and $l+1$ (cf. Fig. 5.2). We aim at deriving an equation

$$\begin{pmatrix} \mathbf{u}_+^{(l)} \\ \mathbf{d}_+^{(l)} \end{pmatrix} = \underline{\mathbf{I}}(l+1,l) \begin{pmatrix} \mathbf{u}_-^{(l+1)} \\ \mathbf{d}_-^{(l+1)} \end{pmatrix} , \tag{5.36}$$

where the *interface matrix* $\underline{\mathbf{I}}(l+1,l)$ connects the phased amplitudes

$$\begin{aligned} \mathbf{u}_+^{(l)} &:= \mathbf{u}^{(l)}(x_l^3) , \\ \mathbf{d}_+^{(l)} &:= \mathbf{d}^{(l)}(x_l^3) , \\ \mathbf{u}_-^{(l+1)} &:= \mathbf{u}^{(l+1)}(x_l^3) , \\ \mathbf{d}_-^{(l+1)} &:= \mathbf{d}^{(l+1)}(x_l^3) , \end{aligned}$$

in both layers at the site of the interface.

By virtue of tangential field vector $\mathbf{V}^{(l)}(x^1,x^2,x^3)$, the continuity conditions of the tangential field components at the interface, derived in Sec. 2.2 and manifested in Eq. (2.19), can be stated as

$$\mathbf{V}^{(l)}(x^1,x^2,x_l^3) \stackrel{!}{=} \mathbf{V}^{(l+1)}(x^1,x^2,x_l^3) \qquad \forall x^1,x^2 . \tag{5.37}$$

Equation (5.37) can be rewritten, using the matrix notation presented in Eq. (5.22), into conditions for the respective amplitudes

$$\underline{\mathbf{M}}^{(l)}(x^1,x^2) \begin{pmatrix} \mathbf{u}_+^{(l)} \\ \mathbf{d}_+^{(l)} \end{pmatrix} \stackrel{!}{=} \underline{\mathbf{M}}^{(l+1)}(x^1,x^2) \begin{pmatrix} \mathbf{u}_-^{(l+1)} \\ \mathbf{d}_-^{(l+1)} \end{pmatrix} \qquad \forall x^1,x^2 . \tag{5.38}$$

The interface matrix defined above is found by multiplication with the inverse of matrix $\underline{\mathbf{M}}^{(l)}$ from the left

$$\begin{pmatrix} \mathbf{u}_+^{(l)} \\ \mathbf{d}_+^{(l)} \end{pmatrix} = \underbrace{\left[\underline{\mathbf{M}}^{(l)}\right]^{-1} \underline{\mathbf{M}}^{(l+1)}}_{\underline{\mathbf{I}}(l+1,l)} \begin{pmatrix} \mathbf{u}_-^{(l+1)} \\ \mathbf{d}_-^{(l+1)} \end{pmatrix} . \tag{5.39}$$

Please note that the inversion of $\underline{\mathbf{M}}^{(l)}$ is meant only symbolic here, since the matrix is still a function of the continuous coordinates x^1 and x^2. In this form it is of size $4 \times 2\bar{N}$ where the rows contain the four transversal field components of the numerical mode. In order to form a proper matrix which can be inverted, we have to discretize these components, e.g., by an expansion into basis functions as in Eq. (5.15). However, there are also other discretization approaches for the purpose of field matching which will be shortly discussed in the subsequent paragraphs. Independent of the concrete matching scheme, the best dimensions for a matrix inversion would be a square matrix of size $2\bar{N} \times 2\bar{N}$, which

means that every numerical mode component must be discretized into $\bar{N}/2 = Ntot$ numbers. In case the matrix cannot be made square, there still exist inversion algorithms like the Moore-Penrose pseudo inverse [77, 78] or the Method of Least Squares [79] that can be applied. For further details we refer the interested reader to the discussion in Ref. [53].

Pointwise Matching

There are several discretization schemes available for the discretization of matrix $\underline{\mathbf{M}}(x^1, x^2)$ so that the fields can be matched at the interface. The first approach is to sample the numerical field components of the eigenmodes contained in $\underline{\mathbf{M}}(x^1, x^2)$ at a finite number $\bar{N}_s$ of test points

$$\mathbf{r}_{\parallel i} = (x_i^1, x_i^2)\,, \qquad i = 1, \ldots, \bar{N}_s\,, \tag{5.40}$$

the same points in both layers. Then the components of the j-th eigenmode become vectors similar to Eq. (5.15), e.g., the H_{2j} component becomes

$$\mathsf{H}_{2j}(\mathbf{r}_\parallel) \quad \rightarrow \quad \Big(\mathsf{H}_{2j}(\mathbf{r}_{\parallel 1}), \ldots, \mathsf{H}_{2j}(\mathbf{r}_{\parallel i}), \ldots, \mathsf{H}_{2j}(\mathbf{r}_{\parallel \bar{N}_s}) \Big)^T\,, \tag{5.41}$$

where the vector entries are not expansion coefficients but field values at the test points.

Consequently, the $2 \times \bar{N}$ submatrices in Eq. (5.24) and the $4 \times \bar{N}$ submatrices in Eq. (5.25) become $2\bar{N}_s \times \bar{N}$ and $4\bar{N}_s \times \bar{N}$ submatrices in sampled form, respectively. The whole sampled eigenmode matrices $\underline{\tilde{\mathbf{M}}}$ for small and large eigenmodes, corresponding to Eq. (5.23) and Eq. (5.25), are then both of dimension $4\bar{N}_s \times 2\bar{N}$ and, hence, the size of the eigenmode matrices is independent of the eigenproblem we use.

Matching by test points holds the advantage that we are not restricted in the choice of basis functions. In particular, it allows for different sets of basis functions in adjacent layers. The downside of pointwise matching is that the sampling points have to be chosen carefully to prevent $\underline{\tilde{\mathbf{M}}}$ from getting singular. Furthermore, the fulfillment of the continuity conditions is solely guaranteed at the sampling points and not in the regions in between [53].

Matching by Overlap Integrals

A second way to match the fields is the matching of eigenmodes of subsequent layers directly by overlap integrals. Overlap integrals determine how much of the energy carried by one eigenmode is transferred to another eigenmode when it is diffracted at the interface. The integration is over the transversal plane, which is why we only need the eigenmodes at the interface as functions of the transversal coordinates $\mathbf{r}_\parallel$. Hence, this procedure is independent of the basis used for the discretization of the eigenmodes as well.

The important part is to find an appropriate overlap integral

$$\mathsf{O}_{jk}^{(l,l+1)} = \int\limits_{\text{Interface}} \mathrm{d}A\; f\Big(\mathsf{E}_{\parallel j}^{(l)}, \mathsf{H}_{\parallel j}^{(l)}, \mathsf{E}_{\parallel k}^{(l+1)}, \mathsf{H}_{\parallel k}^{(l+1)} \Big) \tag{5.42}$$

with a scalar functional f that depends on the eigenmodes, which fulfills the orthonormality condition

$$\mathsf{O}_{jk}^{(l,l)} = \delta_{jk} \tag{5.43}$$

for eigenmodes of the same layer l. This overlap integral clearly defines a scalar product in the infinite vector space of the eigenmodes.

For this overlap integral we can use the orthogonality condition stated in Eqs. (4.26) for waveguide modes derived in Sec. 4.1.2 from the reciprocity theorem for electromagnetic fields. The orthogonality condition is not only valid for waveguide modes, but also for eigenmodes of periodic structures. The overlap integral or its equivalent short hand notation as scalar product then reads

$$\mathsf{O}_{jk}^{(l,l+1)} = \int\limits_{\text{Interface}} \mathrm{d}A \underbrace{\left(\mathbf{E}_j^{(l)} \times \mathbf{H}_k^{(l+1)}\right) \cdot \hat{\mathbf{x}}^3}_{\mathbf{E}_{\parallel j}^{(l)} \cdot \mathbf{H}_{\parallel k}^{(l+1)}} = \left\langle \mathbf{E}_{\parallel j}^{(l)}, \mathbf{H}_{\parallel k}^{(l+1)} \right\rangle . \tag{5.44}$$

This overlap integral is valid for eigenmodes of non-absorbing as well as absorbing structures traveling in the same direction as discussed before. Of course, the eigenmodes must be appropriately normalized.

We illustrate the construction of the interface matrix by means of the large eigensolutions. The field expansion of Eq. (5.20) can be split into two equations where we separate the electric from the magnetic fields. From the continuity condition, Eq. (5.37), we obtain the two relations

$$\sum_{j=1}^{\bar{N}} u_{+,j}^{(l)}\, \mathbf{E}_{\parallel j}^{(l)+} + d_{+,j}^{(l)}\, \mathbf{E}_{\parallel j}^{(l)-} \stackrel{!}{=} \sum_{j=1}^{\bar{N}} u_{-,j}^{(l+1)}\, \mathbf{E}_{\parallel j}^{(l+1)+} + d_{-,j}^{(l+1)}\, \mathbf{E}_{\parallel j}^{(l+1)-} , \tag{5.45a}$$

$$\sum_{j=1}^{\bar{N}} u_{+,j}^{(l)}\, \mathbf{H}_{\parallel j}^{(l)+} + d_{+,j}^{(l)}\, \mathbf{H}_{\parallel j}^{(l)-} \stackrel{!}{=} \sum_{j=1}^{\bar{N}} u_{-,j}^{(l+1)}\, \mathbf{H}_{\parallel j}^{(l+1)+} + d_{-,j}^{(l+1)}\, \mathbf{H}_{\parallel j}^{(l+1)-} . \tag{5.45b}$$

Next, we construct the overlap integrals $\langle\, \cdot\, , \mathbf{H}_{\parallel k}^{(l)+} \rangle$ of Eq. (5.45a), and $\langle \mathbf{E}_{\parallel k}^{(l)+},\, \cdot\, \rangle$ of Eq. (5.45b). We note that, according to Eqs. (4.24),

$$\left\langle \mathbf{E}_{\parallel j}^{(a)-}, \mathbf{H}_{\parallel k}^{(b)+} \right\rangle = \left\langle \mathbf{E}_{\parallel j}^{(a)+}, \mathbf{H}_{\parallel k}^{(b)+} \right\rangle , \tag{5.46a}$$

$$\left\langle \mathbf{E}_{\parallel j}^{(a)+}, \mathbf{H}_{\parallel k}^{(b)-} \right\rangle = -\left\langle \mathbf{E}_{\parallel j}^{(a)+}, \mathbf{H}_{\parallel k}^{(b)+} \right\rangle , \tag{5.46b}$$

hold for modes of arbitrary layers a and b. Using the orthogonality condition Eq. (5.43) and the overlap integral relations Eqs. (5.46), we obtain

$$u_{-,k}^{(l)} + d_{-,k}^{(l)} = \sum_{j=1}^{\bar{N}} \left[u_{+,j}^{(l+1)} + d_{+,j}^{(l+1)} \right] \left\langle \mathbf{E}_{\parallel j}^{(l+1)+}, \mathbf{H}_{\parallel k}^{(l)+} \right\rangle , \tag{5.47a}$$

$$u_{-,k}^{(l)} - d_{-,k}^{(l)} = \sum_{j=1}^{\bar{N}} \left[u_{+,j}^{(l+1)} - d_{+,j}^{(l+1)} \right] \left\langle \mathbf{E}_{\parallel k}^{(l)+}, \mathbf{H}_{\parallel j}^{(l+1)+} \right\rangle . \tag{5.47b}$$

In vector notation this reads

$$\mathbf{u}_-^{(l)} + \mathbf{d}_-^{(l)} = \left(\underline{\mathbf{O}}^{(l+1,l)}\right)^T \left[\mathbf{u}_+^{(l+1)} + \mathbf{d}_+^{(l+1)} \right] , \tag{5.48a}$$

$$\mathbf{u}_-^{(l)} - \mathbf{d}_-^{(l)} = \underline{\mathbf{O}}^{(l,l+1)} \left[\mathbf{u}_+^{(l+1)} - \mathbf{d}_+^{(l+1)} \right] , \tag{5.48b}$$

where we have introduced the overlap matrices $\underline{\mathbf{O}}$ whose elements are defined by Eq. (5.44) incorporating forward modes only.

The interface matrix follows when we add and subtract Eqs. (5.48) as

$$\begin{pmatrix} \mathbf{u}_-^{(l)} \\ \mathbf{d}_-^{(l)} \end{pmatrix} = \underbrace{\begin{pmatrix} \frac{1}{2}\left[\left(\underline{\mathbf{O}}^{(l+1,l)}\right)^T + \underline{\mathbf{O}}^{(l,l+1)}\right] & \frac{1}{2}\left[\left(\underline{\mathbf{O}}^{(l+1,l)}\right)^T - \underline{\mathbf{O}}^{(l,l+1)}\right] \\ \frac{1}{2}\left[\left(\underline{\mathbf{O}}^{(l+1,l)}\right)^T - \underline{\mathbf{O}}^{(l,l+1)}\right] & \frac{1}{2}\left[\left(\underline{\mathbf{O}}^{(l+1,l)}\right)^T + \underline{\mathbf{O}}^{(l,l+1)}\right] \end{pmatrix}}_{\underline{\mathbf{I}}^{(l+1,l)}} \begin{pmatrix} \mathbf{u}_+^{(l+1)} \\ \mathbf{d}_+^{(l+1)} \end{pmatrix} . \quad (5.49)$$

This result is valid for the eigenmodes of the large eigenproblem, when each forward eigenmode is countered by an equivalent backward mode, as well as for the eigenmodes of the small eigenproblem.

Despite our notation, we would like to mention that a different ordering of forward and backward modes is not an issue in practice. It would simply lead to a swapping of rows and columns in the matrices, but it would not change the result.

Matching of Basis Functions

There is a third scheme for the field matching at layer interfaces. Matching by basis functions can be used when the fields in every layer are expanded into the same set of basis functions. The great advantage of this approach is that the fields then match at every point of the interface and not only at some sampling points. However, using the same set of basis functions in every layer imposes a huge restriction. We are particularly interested in this approach since it is used in the Fourier modal method, where the basis functions are plane waves. The Fourier modal method will be introduced in detail in Chap. 6.

In order to come up with an interface matrix, we first recall the expansion of fields into an arbitrary set of basis functions $\mathcal{B}_m(x^1, x^2)$ introduced in Eq. (5.14). We restate the expansion for the ρ-th electric field component of eigenmode j in the l-th layer for convenience:

$$\mathsf{E}_{\rho,j}^{(l)}(x^1, x^2) = \sum_{m=1}^{M} \tilde{\mathsf{E}}_{\rho,mj}^{(l)} \, \mathcal{B}_m(x^1, x^2) \, . \quad (5.50)$$

By definition, the basis functions are the same in each layer and, thus, do not carry a layer label. Furthermore, we leave them in the same order in each layer. Then the field expansion coefficients $\tilde{\mathsf{E}}_{\rho,mj}$ can be seen as elements of an $M \times \bar{N}$ matrix $\underline{\tilde{\mathbf{E}}}_\rho^{(l)}$. These matrices, and the corresponding $\underline{\tilde{\mathbf{H}}}_\rho^{(l)}$ matrices, are the constituents of the mode submatrices $\underline{\tilde{\mathbf{M}}}_\mathsf{E}^{(l)}$, $\underline{\tilde{\mathbf{M}}}_\mathsf{H}^{(l)}$, and $\underline{\tilde{\mathbf{V}}}^{(l)\pm}$, which are the discretized correspondents to the matrices given in Eq. (5.24) and Eq. (5.26), respectively. For example, we find

$$\underline{\tilde{\mathbf{M}}}_\mathsf{H}^{(l)} = \begin{pmatrix} \underline{\tilde{\mathbf{H}}}_1^{(l)} \\ \underline{\tilde{\mathbf{H}}}_2^{(l)} \end{pmatrix} , \qquad \underline{\tilde{\mathbf{V}}}^{(l)\pm} = \begin{pmatrix} -\underline{\tilde{\mathbf{E}}}_2^{(l)} \\ \underline{\tilde{\mathbf{E}}}_1^{(l)} \\ \underline{\tilde{\mathbf{H}}}_1^{(l)} \\ \underline{\tilde{\mathbf{H}}}_2^{(l)} \end{pmatrix} . \quad (5.51)$$

The former matrices are of dimension $2M \times \bar{N}$, the latter of dimension $4M \times \bar{N}$. The mode matrices $\underline{\mathbf{M}}^{(l)}$ are obtained from Eq. (5.23) or Eq. (5.25), and the interface matrix from Eq. (5.39). They are all of dimension $4M \times 2\bar{N}$ and, thus, because $M = \bar{N}/2$ as stipulated in Sec. 5.2.1 in the paragraph about numerical discretization, they are square.

We can use either of these three schemes to connect the layer's fields and construct an interface matrix. Each scheme has its own advantages and disadvantages. The interface matrices are at the core of the scattering matrix algorithm. With their help we can define the scattering matrix iteration in the next section, which specifies the procedure to add an additional layer to an existing S-matrix.

5.3.2. S-Matrix Recursion

The building of scattering matrices is a recursive procedure [24, 76, 80]. From a given scattering matrix connecting layers m through l, $l > m$, we construct the scattering matrix connecting layers m through $l + 1$

$$\underline{\mathbf{S}}(m, l) \quad \xrightarrow{\text{recursion}} \quad \underline{\mathbf{S}}(m, l+1) \tag{5.52}$$

by "adding" the layer $l + 1$.

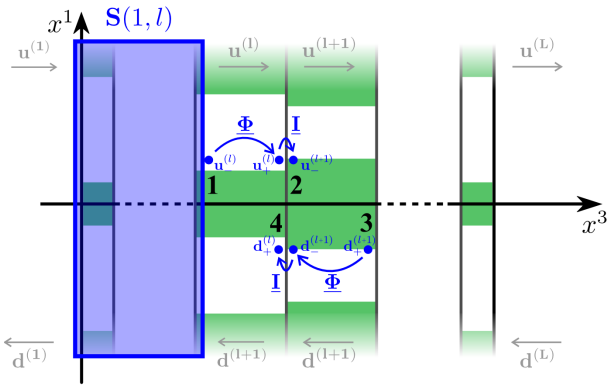


Figure 5.4.: *Schematic illustration of amplitude propagation in one recursion step.*

To this end, we start from the scattering matrix of a single layer without interface which is given by a unit matrix $\underline{\mathbf{S}}(m, m) = \underline{\mathbf{1}}$. The S-matrix $\underline{\mathbf{S}}(m, l)$ is constructed by recursively adding a layer at a time. In terms of Fig. 5.4, one recursion step relates the forward amplitudes $\mathbf{u}$ at points 1 and 2

$$\mathbf{u}_-^{(l)} \quad \xrightarrow{\underline{\mathbf{\Phi}}^{(l)+}(x_l^3)} \quad \mathbf{u}_+^{(l)} \quad \xrightarrow{\underline{\mathbf{I}}(l,l+1)} \quad \mathbf{u}_-^{(l+1)}\,, \tag{5.53a}$$

and the backward amplitudes $\mathbf{d}$ at points 3 and 4

$$\mathbf{d}_+^{(l+1)} \quad \xrightarrow{\underline{\mathbf{\Phi}}^{(l+1)-}(x_l^3)} \quad \mathbf{d}_-^{(l+1)} \quad \xrightarrow{\underline{\mathbf{I}}(l,l+1)} \quad \mathbf{d}_+^{(l)}\,, \tag{5.53b}$$

propagating each phase matrix (entry) in the direction of its decay.

To formalize this recursion step, we recall the scattering matrix definition, Eq. (5.30), and write for an S-matrix connecting layers m and l

$$\begin{pmatrix} \mathbf{u}_l \\ \mathbf{d}_m \end{pmatrix} = \underline{\mathbf{S}}(m,l) \begin{pmatrix} \mathbf{u}_m \\ \mathbf{d}_l \end{pmatrix} = \begin{pmatrix} \underline{\mathbf{S}}_{11} & \underline{\mathbf{S}}_{12} \\ \underline{\mathbf{S}}_{21} & \underline{\mathbf{S}}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{u}_m \\ \mathbf{d}_l \end{pmatrix}, \tag{5.54}$$

where we change the notation and put the layer indices of amplitude vectors into subscripts to improve readability. This means forward amplitudes $\mathbf{u}_l = \mathbf{u}_-^{(l)} = \mathbf{u}^{(l)}$ are always taken at the backward interface (x_{l-1}^3) and backward amplitudes $\mathbf{d}_l = \mathbf{d}_+^{(l)} = \mathbf{d}^{(l)}$ at the forward interface (x_l^3) as originally defined in the field expansion, Sec. 5.2.2. Furthermore, we adapt the phase matrix notation $\underline{\mathbf{\Phi}}_l^+ = \underline{\mathbf{\Phi}}^{(l)+}(x_l^3)$ and $\underline{\mathbf{\Phi}}_{l+1}^- = \underline{\mathbf{\Phi}}^{(l+1)-}(x_l^3)$, which means that the phase matrices are propagated for the whole thicknesses of layers l and $l+1$, respectively. Consequently, the interface matrix defined in Eq. (5.36) reads

$$\begin{pmatrix} \underline{\mathbf{\Phi}}_l^+ \, \mathbf{u}_l \\ \mathbf{d}_l \end{pmatrix} = \underline{\mathbf{I}}(l+1,l) \begin{pmatrix} \mathbf{u}_{l+1} \\ \underline{\mathbf{\Phi}}_{l+1}^- \, \mathbf{d}_{l+1} \end{pmatrix} \tag{5.55}$$

$$= \begin{pmatrix} \underline{\mathbf{I}}_{11} & \underline{\mathbf{I}}_{12} \\ \underline{\mathbf{I}}_{21} & \underline{\mathbf{I}}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{u}_{l+1} \\ \underline{\mathbf{\Phi}}_{l+1}^- \, \mathbf{d}_{l+1} \end{pmatrix}. \tag{5.56}$$

This can be rewritten as

$$\mathbf{u}_l = \underline{\mathbf{S}}_{11}\mathbf{u}_m + \underline{\mathbf{S}}_{12}\mathbf{d}_l \,, \tag{5.57a}$$

$$\mathbf{d}_m = \underline{\mathbf{S}}_{21}\mathbf{u}_m + \underline{\mathbf{S}}_{22}\mathbf{d}_l \,, \tag{5.57b}$$

$$\mathbf{u}_l = \left(\underline{\mathbf{\Phi}}_l^+\right)^{-1} \left[\underline{\mathbf{I}}_{11}\mathbf{u}_{l+1} + \underline{\mathbf{I}}_{12}\underline{\mathbf{\Phi}}_{l+1}^-\mathbf{d}_{l+1}\right] , \tag{5.57c}$$

$$\mathbf{d}_l = \left[\underline{\mathbf{I}}_{21}\mathbf{u}_{l+1} + \underline{\mathbf{I}}_{22}\underline{\mathbf{\Phi}}_{l+1}^-\mathbf{d}_{l+1}\right] . \tag{5.57d}$$

The substitution of Eq. (5.57c) and Eq. (5.57d) into Eq. (5.57a) and solving for $\mathbf{u}_{l+1}$ leads to

$$\begin{aligned} \mathbf{u}_{l+1} &= \left(\underline{\mathbf{I}}_{11} - \underline{\mathbf{\Phi}}_l^+\underline{\mathbf{S}}_{12}\underline{\mathbf{I}}_{21}\right)^{-1} \left[\underline{\mathbf{\Phi}}_l^+\underline{\mathbf{S}}_{11}\ \mathbf{u}_m \ +\ \left(\underline{\mathbf{\Phi}}_l^+\underline{\mathbf{S}}_{12}\underline{\mathbf{I}}_{22} - \underline{\mathbf{I}}_{12}\right)\underline{\mathbf{\Phi}}_{l+1}^-\ \mathbf{d}_{l+1}\right] \\ &= \hat{\underline{\mathbf{S}}}_{11}\ \mathbf{u}_m \ +\ \hat{\underline{\mathbf{S}}}_{12}\ \mathbf{d}_{l+1} \,, \end{aligned} \tag{5.58}$$

from which we can identify $\hat{\underline{\mathbf{S}}}_{11}$ and $\hat{\underline{\mathbf{S}}}_{12}$ as elements of the scattering matrix $\underline{\mathbf{S}}(m,l+1)$. From the substitution of Eq. (5.57d) and Eq. (5.58) into Eq. (5.57b) we obtain

$$\begin{aligned} \mathbf{d}_m &= \left[\underline{\mathbf{S}}_{21} + \underline{\mathbf{S}}_{22}\underline{\mathbf{I}}_{21}\hat{\underline{\mathbf{S}}}_{11}\right]\ \mathbf{u}_m \ +\ \left[\underline{\mathbf{S}}_{22}\underline{\mathbf{I}}_{22}\underline{\mathbf{\Phi}}_{l+1}^- + \underline{\mathbf{S}}_{22}\underline{\mathbf{I}}_{21}\hat{\underline{\mathbf{S}}}_{12}\right]\ \mathbf{d}_{l+1} \\ &= \hat{\underline{\mathbf{S}}}_{21}\ \mathbf{u}_m \ +\ \hat{\underline{\mathbf{S}}}_{22}\ \mathbf{d}_{l+1} \,, \end{aligned} \tag{5.59}$$

from which we can identify the remaining elements $\hat{\underline{\mathbf{S}}}_{21}$ and $\hat{\underline{\mathbf{S}}}_{22}$.

In summary, the recursion formulas for the S-matrix algorithm read

$$\hat{\underline{\mathbf{S}}}_{11} = \left(\underline{\mathbf{I}}_{11} - \underline{\mathbf{\Phi}}_l^+ \underline{\mathbf{S}}_{12}\underline{\mathbf{I}}_{21}\right)^{-1} \underline{\mathbf{\Phi}}_l^+ \underline{\mathbf{S}}_{11}\,, \tag{5.60a}$$

$$\hat{\underline{\mathbf{S}}}_{12} = \left(\underline{\mathbf{I}}_{11} - \underline{\mathbf{\Phi}}_l^+ \underline{\mathbf{S}}_{12}\underline{\mathbf{I}}_{21}\right)^{-1} \left(\underline{\mathbf{\Phi}}_l^+ \underline{\mathbf{S}}_{12}\underline{\mathbf{I}}_{22} - \underline{\mathbf{I}}_{12}\right) \underline{\mathbf{\Phi}}_{l+1}^-\,, \tag{5.60b}$$

$$\hat{\underline{\mathbf{S}}}_{21} = \left[\underline{\mathbf{S}}_{21} + \underline{\mathbf{S}}_{22}\underline{\mathbf{I}}_{21}\hat{\underline{\mathbf{S}}}_{11}\right]\,, \tag{5.60c}$$

$$\hat{\underline{\mathbf{S}}}_{22} = \left[\underline{\mathbf{S}}_{22}\underline{\mathbf{I}}_{22}\underline{\mathbf{\Phi}}_{l+1}^- + \underline{\mathbf{S}}_{22}\underline{\mathbf{I}}_{21}\hat{\underline{\mathbf{S}}}_{12}\right]\,, \tag{5.60d}$$

where $\hat{\underline{\mathbf{S}}}_{ij}$ denotes the elements of scattering matrix $\underline{\mathbf{S}}(m, l+1)$.

The described algorithm is implemented in our numerical framework. It is a rather peculiar choice stemming from Whittaker and Culshaw [24]. However, there exist many variations, like placing the amplitudes in different x^3–positions or defining the interface matrix the other way round.

5.3.3. S-matrix Products

Two scattering matrices $\dot{\underline{\mathbf{S}}} := \underline{\mathbf{S}}(a, b)$ and $\ddot{\underline{\mathbf{S}}} := \underline{\mathbf{S}}(b, c)$ which both include a common layer b can be connected to form a single scattering matrix $\underline{\mathbf{S}} := \underline{\mathbf{S}}(a, c)$. This connection is provided by a special product of the two scattering matrices — the $\star$–product [76, 81]

$$\underline{\mathbf{S}}(a, c) = \underline{\mathbf{S}}(a, b) \star \underline{\mathbf{S}}(b, c)\,. \tag{5.61}$$

The $\star$–product can be derived from the defining equations of $\underline{\mathbf{S}}(a, b)$ and $\underline{\mathbf{S}}(b, c)$ by following an approach similar to what we did in Sec. 5.3.2: We decompose them into a set of four equations and, by substituting and reordering, solve them for the amplitudes $\mathbf{u}^{(c)}$ and $\mathbf{d}^{(a)}$ in dependence of $\mathbf{u}^{(a)}$ and $\mathbf{d}^{(c)}$. We skip the detailed derivation and simply state the final result. Thus, the $\star$–product is defined as

$$\underline{\mathbf{S}}_{11} = \ddot{\underline{\mathbf{S}}}_{11}\left[\underline{\mathbf{1}} - \dot{\underline{\mathbf{S}}}_{12}\,\ddot{\underline{\mathbf{S}}}_{21}\right]^{-1} \dot{\underline{\mathbf{S}}}_{11}\,, \tag{5.62a}$$

$$\underline{\mathbf{S}}_{12} = \ddot{\underline{\mathbf{S}}}_{12} + \ddot{\underline{\mathbf{S}}}_{11}\left[\underline{\mathbf{1}} - \dot{\underline{\mathbf{S}}}_{12}\,\ddot{\underline{\mathbf{S}}}_{21}\right]^{-1} \dot{\underline{\mathbf{S}}}_{12}\,\ddot{\underline{\mathbf{S}}}_{22}\,, \tag{5.62b}$$

$$\underline{\mathbf{S}}_{21} = \dot{\underline{\mathbf{S}}}_{21} + \dot{\underline{\mathbf{S}}}_{22}\,\ddot{\underline{\mathbf{S}}}_{21}\left[\underline{\mathbf{1}} - \dot{\underline{\mathbf{S}}}_{12}\,\ddot{\underline{\mathbf{S}}}_{21}\right]^{-1} \dot{\underline{\mathbf{S}}}_{11}\,, \tag{5.62c}$$

$$\underline{\mathbf{S}}_{22} = \dot{\underline{\mathbf{S}}}_{22}\left(\underline{\mathbf{1}} + \ddot{\underline{\mathbf{S}}}_{21}\left[\underline{\mathbf{1}} - \dot{\underline{\mathbf{S}}}_{12}\,\ddot{\underline{\mathbf{S}}}_{21}\right]^{-1} \dot{\underline{\mathbf{S}}}_{11}\right)\ddot{\underline{\mathbf{S}}}_{22}\,. \tag{5.62d}$$

It can be shown that this product is associative. It is valid independent of the variation in the algorithm of how the S-matrix is obtained and can also be used for other matrices (interface-matrix, phase-matrix), as long as they are of S-matrix type, i.e., relate incoming to outgoing amplitudes.

5.3.4. Transformation of a T-Matrix into an S-Matrix

A matrix of T-matrix type (cf. Eq. (5.29)) can be transformed into an S-matrix by a similar procedure as in Sec. 5.3.3 [76]. If the T-matrix connecting layers a and b is given by

$$\begin{pmatrix} \mathbf{u}^{(b)} \\ \mathbf{d}^{(b)} \end{pmatrix} = \begin{pmatrix} \underline{\mathbf{T}}_{11} & \underline{\mathbf{T}}_{12} \\ \underline{\mathbf{T}}_{21} & \underline{\mathbf{T}}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{u}^{(a)} \\ \mathbf{d}^{(a)} \end{pmatrix}\,, \tag{5.63}$$

the corresponding scattering matrix elements are given by

$$\underline{\mathbf{S}}_{11} = \underline{\mathbf{T}}_{11} - \underline{\mathbf{T}}_{12}\,\underline{\mathbf{T}}_{22}^{-1}\,\underline{\mathbf{T}}_{21}\,, \tag{5.64a}$$
$$\underline{\mathbf{S}}_{12} = \underline{\mathbf{T}}_{12}\,\underline{\mathbf{T}}_{22}^{-1}\,, \tag{5.64b}$$
$$\underline{\mathbf{S}}_{21} = -\,\underline{\mathbf{T}}_{22}^{-1}\,\underline{\mathbf{T}}_{21}\,, \tag{5.64c}$$
$$\underline{\mathbf{S}}_{22} = \underline{\mathbf{T}}_{22}^{-1}\,. \tag{5.64d}$$

With the help of this scheme, the interface- and phase-matrices can be rewritten into S-matrices.

5.3.5. Reversion of Layer Sequence

There is an easy way to obtain the scattering matrix $\underline{\mathbf{S}}(l,m)$ from the scattering matrix $\underline{\mathbf{S}}(m,l)$, which can be quite useful to efficiently exploit symmetries in a layer sequence. Looking at Fig. 5.5, we note that the right-hand side system is equivalent to the central system, it is just rotated. If we further swap the amplitude labels $\mathbf{u}_m \leftrightarrow \mathbf{d}_m$ and $\mathbf{d}_l \leftrightarrow \mathbf{u}_l$ so that forward amplitudes become backward amplitudes and vice versa, we get the S-matrix of the reverted layer sequence on the right-hand side.

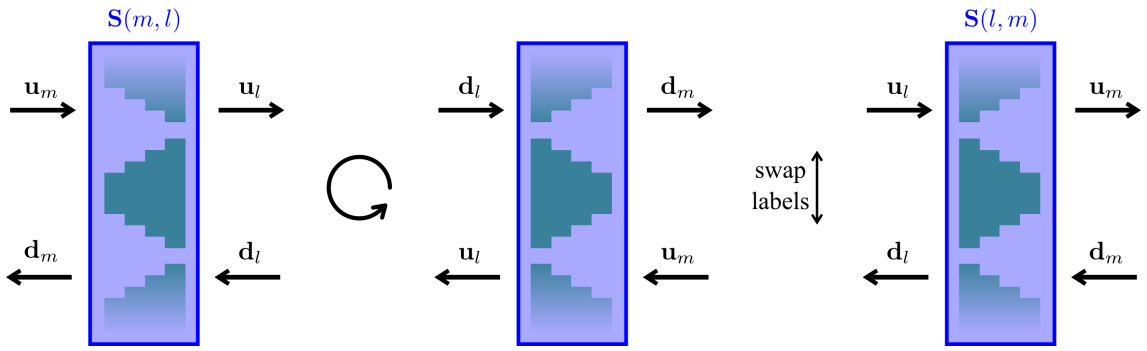


Figure 5.5.: *Sketch: Reversion of layer sequence. Details see text.*

Starting from the defining equation for $\underline{\mathbf{S}}(l,m)$ we derive the defining equation for $\underline{\mathbf{S}}(m,l)$ in two steps. The first step is the relabeling of the amplitudes described above. The second step is a reorder-

ing of rows and columns. Written down as equations this reads

$$
\begin{aligned}
&\begin{pmatrix} \mathbf{u}_l \\ \mathbf{d}_m \end{pmatrix} = \overbrace{\begin{pmatrix} \underline{\mathbf{S}}_{11} & \underline{\mathbf{S}}_{12} \\ \underline{\mathbf{S}}_{21} & \underline{\mathbf{S}}_{22} \end{pmatrix}}^{\underline{\mathbf{S}}(m,l)} \begin{pmatrix} \mathbf{u}_m \\ \mathbf{d}_l \end{pmatrix} , \\
&\mathbf{u}_l \leftrightarrow \mathbf{d}_l \quad \Big\Downarrow \quad \mathbf{d}_m \leftrightarrow \mathbf{u}_m \\
&\begin{pmatrix} \mathbf{d}_l \\ \mathbf{u}_m \end{pmatrix} = \begin{pmatrix} \underline{\mathbf{S}}_{11} & \underline{\mathbf{S}}_{12} \\ \underline{\mathbf{S}}_{21} & \underline{\mathbf{S}}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{d}_m \\ \mathbf{u}_l \end{pmatrix} , \\
&\qquad \Big\Downarrow \quad \text{reorder} \\
&\begin{pmatrix} \mathbf{u}_m \\ \mathbf{d}_l \end{pmatrix} = \underbrace{\begin{pmatrix} \underline{\mathbf{S}}_{22} & \underline{\mathbf{S}}_{21} \\ \underline{\mathbf{S}}_{12} & \underline{\mathbf{S}}_{11} \end{pmatrix}}_{\underline{\mathbf{S}}(l,m)} \begin{pmatrix} \mathbf{u}_l \\ \mathbf{d}_m \end{pmatrix} .
\end{aligned}
\tag{5.65}
$$

We find that the scattering matrix for the reversed layer sequence is given by swapping the row and column index of the submatrices.

Please note that a reversion of the layer sequence in this easy way is only possible because of our peculiar symmetric choice of the amplitude positions. Thus, the reversion might not work with other S-matrix schemes if the propagation through the layer is not applied in a symmetric order.

5.3.6. Reduction to Half Size or Fast Scattering Matrix Algorithm

From Eqs. (5.35) we have learned that the transmitted and reflected amplitudes can be obtained from submatrices $\underline{\mathbf{S}}_{11}$ and $\underline{\mathbf{S}}_{21}$ of the system scattering matrix $\underline{\mathbf{S}}(1, L)$ provided the illumination is from the front, i.e., ($\mathbf{u}^{(1)} \neq \mathbf{0}$, $\mathbf{d}^{(L)} = \mathbf{0}$).

If we are only interested in the transmitted and reflected amplitudes of the whole system and not in the amplitudes in the intermediate layers, there is a scattering matrix scheme which only needs two of the four submatrices of the scattering matrix and, thus, enormously speeds up the calculation [76]. The trick is to illuminate the structure from the back ($\mathbf{u}^{(1)} = \mathbf{0}$, $\mathbf{d}^{(L)} \neq \mathbf{0}$) instead of from the front. Then, the transmitted and reflected amplitudes

$$
\begin{aligned}
&\text{reflected amplitudes:} && \mathbf{u}^{(L)} = \underline{\mathbf{S}}_{12}(1, L)\, \mathbf{d}^{(L)} , && \text{(5.66a)} \\
&\text{transmitted amplitudes:} && \mathbf{d}^{(1)} = \underline{\mathbf{S}}_{22}(1, L)\, \mathbf{d}^{(L)} . && \text{(5.66b)}
\end{aligned}
$$

are provided by the submatrices $\underline{\mathbf{S}}_{12}$ and $\underline{\mathbf{S}}_{22}$. Note that, in order to retain the same physical situation, we have to construct the scattering matrix with the reversed layer sequence if the sequence is not symmetric anyways.

The advantage of using these submatrices becomes apparent by looking at the S-matrix recursion formulas, Eqs. (5.60). Submatrices $\underline{\mathbf{S}}_{12}$ (cf. Eq. (5.60b)) and $\underline{\mathbf{S}}_{22}$ (cf. Eq. (5.60d)) only depend on the same submatrices of the previous S-matrix or on themselves, i.e., it is sufficient to calculate only these matrices in every layer. Instead, the recursion of $\underline{\mathbf{S}}_{11}$ and $\underline{\mathbf{S}}_{21}$ involves all four submatrices.

The drawback of this so-called *fast* scattering matrix algorithm is that we need all submatrices to calculate the amplitudes in intermediate layers (cf. Eqs. (5.68)). Hence, it is not possible to calculate the fields or any related quantities in layers other than the first and last ones when we use the fast algorithm.

5.3.7. Expansion Amplitudes in Arbitrary Layers

The scattering matrix of the whole system $\underline{\mathbf{S}}(1, L)$ relates the amplitudes in the first and last layers only. If we are interested in the field distribution of an intermediate layer l, we cannot obtain the necessary amplitudes $\mathbf{u}^{(l)}$ and $\mathbf{d}^{(l)}$ from this S-matrix. Instead, we need the two matrices $\underline{\mathbf{S}}(1, l)$ and $\underline{\mathbf{S}}(l, L)$. We need both because the outgoing and incoming amplitudes in layer l of the first matrix are the incoming and outgoing amplitudes of the second one, and vice versa. Both scattering matrices are coupled via the amplitudes in layer l. It is the nature of multiple diffractions at the layer interfaces that are described by the scattering matrices.

Hence, we note down the defining equations for the two S-matrices

$$\begin{aligned}\begin{pmatrix}\mathbf{u}_l\\ \mathbf{d}_1\end{pmatrix} &= \underline{\mathbf{S}}(1,l)\begin{pmatrix}\mathbf{u}_1\\ \mathbf{d}_l\end{pmatrix}\\ &= \begin{pmatrix}\underline{\mathbf{S}}_{11} & \underline{\mathbf{S}}_{12}\\ \underline{\mathbf{S}}_{21} & \underline{\mathbf{S}}_{22}\end{pmatrix}\begin{pmatrix}\mathbf{u}_1\\ \mathbf{d}_l\end{pmatrix},\end{aligned} \tag{5.67a}$$

and

$$\begin{aligned}\begin{pmatrix}\mathbf{u}_L\\ \mathbf{d}_l\end{pmatrix} &= \underline{\mathbf{S}}(l,L)\begin{pmatrix}\mathbf{u}_l\\ \mathbf{d}_L\end{pmatrix}\\ &= \begin{pmatrix}\hat{\underline{\mathbf{S}}}_{11} & \hat{\underline{\mathbf{S}}}_{12}\\ \hat{\underline{\mathbf{S}}}_{21} & \hat{\underline{\mathbf{S}}}_{22}\end{pmatrix}\begin{pmatrix}\mathbf{u}_l\\ \mathbf{d}_L\end{pmatrix}.\end{aligned} \tag{5.67b}$$

We solve the contained four equations for the amplitudes in layer l in dependence of the input amplitudes in layers 1 and L, namely $\mathbf{u}_1$ and $\mathbf{d}_L$, by substitution and reordering. The amplitudes in question are then given as

$$\mathbf{u}_l = \left[\left(\underline{\mathbf{1}} - \underline{\mathbf{S}}_{12}\,\hat{\underline{\mathbf{S}}}_{21}\right)^{-1}\underline{\mathbf{S}}_{11}\right]\mathbf{u}_1 + \left[\left(\underline{\mathbf{1}} - \underline{\mathbf{S}}_{12}\,\hat{\underline{\mathbf{S}}}_{21}\right)^{-1}\underline{\mathbf{S}}_{12}\,\hat{\underline{\mathbf{S}}}_{22}\right]\mathbf{d}_L\,, \tag{5.68a}$$

$$\mathbf{d}_l = \left[\left(\underline{\mathbf{1}} - \hat{\underline{\mathbf{S}}}_{21}\,\underline{\mathbf{S}}_{12}\right)^{-1}\hat{\underline{\mathbf{S}}}_{21}\,\underline{\mathbf{S}}_{12}\right]\mathbf{u}_1 + \left[\left(\underline{\mathbf{1}} - \hat{\underline{\mathbf{S}}}_{21}\,\underline{\mathbf{S}}_{12}\right)^{-1}\hat{\underline{\mathbf{S}}}_{22}\right]\mathbf{d}_L\,. \tag{5.68b}$$

We note that if the fields are required in every layer, the scattering matrix algorithm is quite inefficient: Though we can successively "add" a layer $(l+1)$ to the matrix $\underline{\mathbf{S}}(1, l)$, we are not aware

of any numerically stable way to easily "remove" a layer of matrix $\underline{\mathbf{S}}(l, L)$ to obtain $\underline{\mathbf{S}}(l+1, L)$. Therefore, the matrix $\underline{\mathbf{S}}(l+1, L)$ must be built from scratch for each single layer, which makes the whole procedure rather expensive.

Here, we conclude the discussion of the scattering matrix algorithm and modal methods in general. In Chap. 6 we transfer these findings into one concrete modal method implementation — the Fourier modal method.

Chapter 6.
6 Fourier Modal Method

The Fourier modal method is one particular simulation method for fully vectorial, three-dimensional light propagation according to Maxwell's equations which fits into the scheme of modal methods introduced in Chap. 5. The method aims at micro- and nano-scale structures with transversal periodicity in one or two dimensions such as amongst others photonic crystals (PC) and meta-materials. Therefore, it builds upon plane waves as expansion basis which is well suited for this purpose because the basis functions feature a similar periodicity as the Bloch-periodic eigenmodes of the system. This guarantees an efficient representation. Naturally, with this basis choice, the method is not equally well suited for the description of local effects with small transversal extent compared to the size of one structure period.

Like all modal methods, the FMM relies on layered structures, and introduces, where not already existant, an artificial decomposition into slices by a staircase approximation. Each periodic layer can be considered as a diffraction grating, which is well known from elementary physics, and the entire system as a stack of such subsequent gratings. Hence, electromagnetic waves incident onto the system are either reflected or transmitted into Bragg orders, and the method is attributed to the field of diffractive optics.

Within the layer, the structure and field description with plane waves is achieved by Fourier expansion. The Fast Fourier Transformation (FFT) algorithm is employed to enable an efficient calculation of the Fourier coefficients. This requires a real-space discretization of the unit cell on a Cartesian grid. The matching between adjacent layers in the scattering matrix algorithm is achieved by basis functions. What is true for the modal methods is also true for the FMM: with the expansion into eigenmodes and the s-matrix algorithm, it is very efficient in the handling of uniform structures and repeating patterns along the main propagation direction — or any combination thereof.

This chapter picks up many topics of Chap. 5 again — like basis functions (Sec. 6.1), field expansion and discretization (Sec. 6.2), eigenproblem (Sec. 6.3), sources (Sec. 6.5), and output quantities (Sec. 6.7 and Sec. 6.6) — but this time in a concrete, FMM-specific way. The aim of this high degree of detailedness is to provide the reader with a full insight into the method, its capabilities, and implementation peculiarities. Furthermore, in Sec. 6.8 we introduce the theoretical background for and a concrete example of symmetry reduction within the FMM which is handy in order to lower the computational costs for the simulation of appropriate systems.

6.1. Plane Wave Basis

The Fourier modal method builds upon plane wave basis functions

$$\mathcal{B}_m(\mathbf{r}_\parallel) = e^{\mathrm{i}\mathbf{k}_{\parallel,m}\cdot\mathbf{r}_\parallel} = e^{\mathrm{i}(\alpha_{m_1}x^1+\beta_{m_2}x^2)} = \mathcal{B}_{m_1}(x^1)\cdot\mathcal{B}_{m_2}(x^2) \tag{6.1}$$

in the two transverse dimensions. The multi-index $m = (m_1, m_2)$ labels a reciprocal lattice point and has been defined in Chap. 3. The wave vector's transverse part, which is parallel to the layer interfaces, is denoted with $\mathbf{k}_{\parallel,m}$ and its components have been defined in Eqs. (3.47). The set of all basis functions $\{\mathcal{B}_m\}$ is called the *plane wave basis*. The basis functions include the wave vector's parallel part $\mathbf{k}_\parallel = (\alpha_0, \beta_0)^T$ which is inherited from the incident wave.

The plane wave basis is orthonormal with respect to the scalar product

$$\langle\mathcal{B}_m, \mathcal{B}_n\rangle = \frac{1}{a_1a_2}\int_0^{a_1}\mathrm{d}x^1\int_0^{a_2}\mathrm{d}x^2\;\mathcal{B}_m^*(x^1,x^2)\,\mathcal{B}_n(x^1,x^2) = \delta_{m_1n_1}\delta_{m_2n_2} = \delta_{mn} \tag{6.2}$$

where the integral is taken over the dimensions of the unit cell. Furthermore, the infinite plane wave basis is complete

$$\delta(x^1,x^2) = \frac{1}{a_1a_2}\sum_{m=0}^{\infty}\mathcal{B}_m(x^1,x^2)\,, \tag{6.3}$$

which guarantees that every continuous function can be represented in the infinite plane wave basis. As alluded to in Chap. 3, the Fourier series expansion must be truncated for a numerical treatment. Then, Eq. (6.3) is only approximately true.

6.2. Field Expansion and Discretization of Operators

In Sec. 5.2.1 we derived the full anisotropic eigenvalue equations in direct space, Eq. (5.10). The task at hand is to discretize this equation with the help of the plane wave basis and transform the problem into reciprocal space. The derivation of the small eigenproblem in Fourier space can be achieved similarly.

With the Floquet-Fourier expansion of the Field components, Eq. (3.15), and the $\hat{\mathrm{L}}_\rho$ operators applied to the material tensors, which we have introduced in Eq. (3.38), this task is easily achieved. The former, exemplarily written out for the first component of the electric field, reads

$$E_1(x^1,x^2) \approx \mathsf{E}_1(x^1,x^2) = \sum_{m=1}^{M}\tilde{\mathsf{E}}_{1,m}\,e^{\mathrm{i}(\alpha_{m_1}x^1+\beta_{m_2}x^2)}\,. \tag{6.4}$$

Recalling the general discretization procedure described in Sec. 5.2.1, we represent the Floquet-Fourier series of the field component by its coefficient vector

$$\tilde{\mathbf{\mathsf{E}}}_1 = (\tilde{\mathsf{E}}_{1,1},\ldots,\tilde{\mathsf{E}}_{1,m},\ldots,\tilde{\mathsf{E}}_{1,M})^T\,. \tag{6.5}$$

Having the fields expanded, the derivatives in operator $\underline{\boldsymbol{\mathcal{A}}}$ (cf. Eq. (5.8)) can be carried out which leads to the substitution

$$\frac{\partial_1}{\mathrm{i}} \to \alpha_{m_1}\,, \quad \text{and} \quad \frac{\partial_2}{\mathrm{i}} \to \beta_{m_2}\,, \tag{6.6}$$

for each expansion term separately. This is not quite obvious for the derivatives which are left of permittivity or permeability tensor components, because these depend on the spatial coordinates themselves. However, considering the derivation of the eigenvalue equation, i.e., Eq. (B.5), Eq. (B.6), Eq. (B.7), and Eq. (B.8), we notice that the derivatives originally stand in front of the third field components E_3 and H_3 which were occasionally eliminated. But the third components have an equivalent Floquet-Fourier expansion as the other field components. Thus, if we take the derivatives before the elimination it becomes clear that the proper replacements are indeed as given above. In matrix notation the correct replacement is

$$\frac{\partial_1}{\mathrm{i}} \to \underline{\boldsymbol{\alpha}} \,, \quad \text{and} \quad \frac{\partial_2}{\mathrm{i}} \to \underline{\boldsymbol{\beta}} \,, \tag{6.7}$$

where $\underline{\boldsymbol{\alpha}}$ and $\underline{\boldsymbol{\beta}}$ are diagonal matrices with α_{m_1} and β_{m_2} as their (m,m)-th entry, respectively.

This is also supported by the fact that the first terms in every operator component $\mathcal{A}^{ij}$, $i,j = 1,2,3,4$, are products that obey the factorization rules. Thus, the expansion of those products is (cf. Eq. (3.25))

$$\varepsilon^{\rho\sigma}(x^1,x^2)\, E_\sigma(x^1,x^2) = \sum_{m=1}^{M} \left(\sum_{n=1}^{M} \tilde{\varepsilon}^{\rho\sigma}_{m-n}\, \tilde{E}_{\sigma,n} \right) e^{\mathrm{i}(\alpha_{m_1} x^1 + \beta_{m_2} x^2)} \tag{6.8}$$

and similar for the magnetic terms. In matrix-vector notation the corresponding replacement in operator $\underline{\boldsymbol{\mathcal{A}}}$ is of the form

$$\varepsilon^{\rho\sigma}(x^1,x^2)\,\mathrm{X}\, E_\sigma(x^1,x^2) \quad \to \quad [\![\varepsilon^{\rho\sigma}]\!]\, \underline{\boldsymbol{X}}\, \tilde{\mathbf{E}}_\sigma \,, \tag{6.9}$$

where X is either 1 or $\frac{\partial_\rho}{\mathrm{i}}$, and $\underline{\boldsymbol{X}}$ either the unit matrix $\underline{\mathbf{1}}$ or one of $\underline{\boldsymbol{\alpha}}$ or $\underline{\boldsymbol{\beta}}$, respectively.

We would like to emphasize that the quantities $\check{\varepsilon}^{\rho\sigma}$ and $\check{\mu}^{\rho\sigma}$ are only placeholders for combinations of material tensor components defined by the operator $\hat{\mathrm{l}}_3^-$ which reshuffles the terms in the required way (cf. Eq. (5.6) and Eq. (3.38)). Hence, in reciprocal space, $\check{\varepsilon}^{\rho\sigma}$ and $\check{\mu}^{\rho\sigma}$ are replaced by components of the Toeplitz matrix tensors[1] $\underline{\tilde{\boldsymbol{\varepsilon}}}$ and $\underline{\tilde{\boldsymbol{\mu}}}$, with

$$\underline{\check{\boldsymbol{\varepsilon}}}(x^1,x^2) \quad \to \quad \underline{\hat{\boldsymbol{\varepsilon}}} = \hat{\mathrm{l}}_3^- \, \frac{\hat{\mathrm{L}}_2\hat{\mathrm{L}}_1 + \hat{\mathrm{L}}_1\hat{\mathrm{L}}_2}{2}\, \underline{\boldsymbol{\varepsilon}}(x^1,x^2) \,, \tag{6.10a}$$

and

$$\underline{\check{\boldsymbol{\mu}}}(x^1,x^2) \quad \to \quad \underline{\hat{\boldsymbol{\mu}}} = \hat{\mathrm{l}}_3^- \, \frac{\hat{\mathrm{L}}_2\hat{\mathrm{L}}_1 + \hat{\mathrm{L}}_1\hat{\mathrm{L}}_2}{2}\, \underline{\boldsymbol{\mu}}(x^1,x^2) \,. \tag{6.10b}$$

The average of the two $\hat{\mathrm{L}}_\rho$ operators in different order has been discussed in Sec. 3.4.4.

Note that, for the step of rewriting the whole equation into matrix vector notation, the usual procedure is to implicitly compare coefficients of the same basis functions on the left and the right hand side. This results in the vanishing of the exponential terms.

[1] This means, the components of the tensor are Toeplitz matrices.

6.3. Eigenproblem

Finally, we have all ingredients to write down the discretized eigenvalue equation and system operator of the Fourier modal method. We provide both the large system operator for full anisotropic systems in Sec. 6.3.1, as well as the small system operators for in-plane anisotropy or less complex material tensors in Sec. 6.3.2. To complete the discussion of the eigenproblem solution, we give details about the numerical algorithms used to diagonalize the eigenproblems in Sec. 6.3.3.

6.3.1. Full Anisotropy — Large Eigenproblem

The discretized full anisotropic eigenvalue equation of the Fourier modal method can finally be written as

$$\gamma \begin{pmatrix} -\tilde{\mathbf{E}}_2 \\ \tilde{\mathbf{E}}_1 \\ \tilde{\mathbf{H}}_1 \\ \tilde{\mathbf{H}}_2 \end{pmatrix} = \underline{\mathbf{A}} \begin{pmatrix} -\tilde{\mathbf{E}}_2 \\ \tilde{\mathbf{E}}_1 \\ \tilde{\mathbf{H}}_1 \\ \tilde{\mathbf{H}}_2 \end{pmatrix}, \tag{6.11}$$

where the discretized system matrix is given by

$$\underline{\mathbf{A}} = \begin{pmatrix} -\underline{\boldsymbol{\beta}}\,\underline{\hat{\boldsymbol{\varepsilon}}}^{32} - \underline{\hat{\boldsymbol{\mu}}}^{13}\,\underline{\boldsymbol{\alpha}} & \underline{\boldsymbol{\beta}}\,\underline{\hat{\boldsymbol{\varepsilon}}}^{31} - \underline{\hat{\boldsymbol{\mu}}}^{13}\,\underline{\boldsymbol{\beta}} & \omega^2\underline{\hat{\boldsymbol{\mu}}}^{11} - \underline{\boldsymbol{\beta}}\,\underline{\hat{\boldsymbol{\varepsilon}}}^{33}\,\underline{\boldsymbol{\beta}} & \omega^2\underline{\hat{\boldsymbol{\mu}}}^{12} + \underline{\boldsymbol{\beta}}\,\underline{\hat{\boldsymbol{\varepsilon}}}^{33}\,\underline{\boldsymbol{\alpha}} \\ \underline{\boldsymbol{\alpha}}\,\underline{\hat{\boldsymbol{\varepsilon}}}^{32} - \underline{\hat{\boldsymbol{\mu}}}^{23}\,\underline{\boldsymbol{\alpha}} & -\underline{\boldsymbol{\alpha}}\,\underline{\hat{\boldsymbol{\varepsilon}}}^{31} - \underline{\hat{\boldsymbol{\mu}}}^{23}\,\underline{\boldsymbol{\beta}} & \omega^2\underline{\hat{\boldsymbol{\mu}}}^{21} + \underline{\boldsymbol{\alpha}}\,\underline{\hat{\boldsymbol{\varepsilon}}}^{33}\,\underline{\boldsymbol{\beta}} & \omega^2\underline{\hat{\boldsymbol{\mu}}}^{22} - \underline{\boldsymbol{\alpha}}\,\underline{\hat{\boldsymbol{\varepsilon}}}^{33}\,\underline{\boldsymbol{\alpha}} \\ \underline{\hat{\boldsymbol{\varepsilon}}}^{22} - \frac{1}{\omega^2}\underline{\boldsymbol{\alpha}}\,\underline{\hat{\boldsymbol{\mu}}}^{33}\,\underline{\boldsymbol{\alpha}} & -\underline{\hat{\boldsymbol{\varepsilon}}}^{21} - \frac{1}{\omega^2}\underline{\boldsymbol{\alpha}}\,\underline{\hat{\boldsymbol{\mu}}}^{33}\,\underline{\boldsymbol{\beta}} & -\underline{\hat{\boldsymbol{\varepsilon}}}^{23}\,\underline{\boldsymbol{\beta}} - \underline{\boldsymbol{\alpha}}\,\underline{\hat{\boldsymbol{\mu}}}^{31} & \underline{\hat{\boldsymbol{\varepsilon}}}^{23}\,\underline{\boldsymbol{\alpha}} - \underline{\boldsymbol{\alpha}}\,\underline{\hat{\boldsymbol{\mu}}}^{32} \\ -\underline{\hat{\boldsymbol{\varepsilon}}}^{12} - \frac{1}{\omega^2}\underline{\boldsymbol{\beta}}\,\underline{\hat{\boldsymbol{\mu}}}^{33}\,\underline{\boldsymbol{\alpha}} & \underline{\hat{\boldsymbol{\varepsilon}}}^{11} - \frac{1}{\omega^2}\underline{\boldsymbol{\beta}}\,\underline{\hat{\boldsymbol{\mu}}}^{33}\,\underline{\boldsymbol{\beta}} & \underline{\hat{\boldsymbol{\varepsilon}}}^{13}\,\underline{\boldsymbol{\beta}} - \underline{\boldsymbol{\beta}}\,\underline{\hat{\boldsymbol{\mu}}}^{31} & -\underline{\hat{\boldsymbol{\varepsilon}}}^{13}\,\underline{\boldsymbol{\alpha}} - \underline{\boldsymbol{\beta}}\,\underline{\hat{\boldsymbol{\mu}}}^{32} \end{pmatrix}. \tag{6.12}$$

6.3.2. In-Plane Anisotropy — Small Eigenproblem

The small discretized eigenvalue equations, Eqs. (5.13), which cover in-plane anisotropy of the material tensors, then read

$$\gamma^2 \begin{pmatrix} -\tilde{\mathbf{E}}_2 \\ \tilde{\mathbf{E}}_1 \end{pmatrix} = \underline{\mathbf{F}}\,\underline{\mathbf{G}} \begin{pmatrix} -\tilde{\mathbf{E}}_2 \\ \tilde{\mathbf{E}}_1 \end{pmatrix}, \tag{6.13a}$$

$$\gamma^2 \begin{pmatrix} \tilde{\mathbf{H}}_1 \\ \tilde{\mathbf{H}}_2 \end{pmatrix} = \underline{\mathbf{G}}\,\underline{\mathbf{F}} \begin{pmatrix} \tilde{\mathbf{H}}_1 \\ \tilde{\mathbf{H}}_2 \end{pmatrix}. \tag{6.13b}$$

with discretized operators $\underline{\mathbf{F}}$ and $\underline{\mathbf{G}}$ obtained from the submatrices of operator $\underline{\mathbf{A}}$ as defined in Eq. (5.9).

Diagonal material tensors and isotropic material tensors reduce the complexity of the system operators $\underline{\mathbf{F}}$ and $\underline{\mathbf{G}}$ even further. The dedicated eigenproblems are straight forwardly deduced by setting the respective components in Eq. (6.12) to zero. Appendix B.1 lists some more of those system operators as implemented in our code.

6.3.3. Numerical Solution

At this point, we would like to emphasize that our numerical implementation uses the lattice constant a_1 as normalization constant (cf. scaling factor a in Sec. 2.1.5).

The discretized eigenvalue equation is numerically solved with the routine *zgeev* from the Fortran library LAPACK (Linear Algebra PACKage) [82], or the respective MKL (Math Kernel Library) implementation [42] on Intel machines. This routine calculates the eigensolutions of non-symmetric complex eigenproblems in double precision. The eigenproblem must be solved for each distinguishable layer separately.

In case of the large eigenproblem, where $\underline{\mathbf{A}}$ is a square matrix of dimension $4M \times 4M$, the output of the routine for the eigenproblem of layer l is the discretized eigenmode matrix $\underline{\tilde{\mathbf{M}}}^{(l)}$ corresponding to Eq. (5.25), and a vector containing the associated eigenvalues γ_j. The output has to be sorted and ordered into forward and backward modes resulting in the eigenmode matrices $\underline{\tilde{\mathbf{V}}}^{(l)\pm}$ (cf. Eq. (5.26)), and the eigenvalues $\gamma_j^{(l)\pm}$.

The output of the small discretized eigenproblem in layer l, where the combined matrix operators give a square matrix of dimension $2M \times 2M$, is either the eigenmode matrix $\underline{\tilde{\mathbf{M}}}_{\mathsf{E}}^{(l)}$ for Eq. (6.13a), or $\underline{\tilde{\mathbf{M}}}_{\mathsf{H}}^{(l)}$ for Eq. (6.13b) in analogy to the real-space matrices defined in equation Eq. (5.24). The missing submatrix is then obtained from

$$\underline{\tilde{\mathbf{M}}}_{\mathsf{H}}^{(l)} = \underline{\gamma}^{(l)} \underline{\mathbf{G}}^{(l)} \underline{\tilde{\mathbf{M}}}_{\mathsf{E}}^{(l)} , \tag{6.14a}$$

or

$$\underline{\tilde{\mathbf{M}}}_{\mathsf{E}}^{(l)} = \underline{\gamma}^{(l)} \underline{\mathbf{F}}^{(l)} \underline{\tilde{\mathbf{M}}}_{\mathsf{H}}^{(l)} , \tag{6.14b}$$

respectively, where $\underline{\gamma} = \mathrm{diag}[\gamma_1, \ldots, \gamma_M]$ (cf. Eq. (5.19)). The discretized eigenmode matrix $\underline{\tilde{\mathbf{M}}}^{(l)}$ is then obtained similar to Eq. (5.23).

Consequently, the eigenmode matrix $\underline{\tilde{\mathbf{M}}}^{(l)}$, which contains the expansion coefficients of all eigenmodes as column vectors, is of size $4M \times 2\bar{N}$. M is the number of retained plane waves, which we choose equal for each field component. Furthermore, we take the same selection of reciprocal lattice vectors for each expansion. $2\bar{N}$ is the total number of eigensolutions provided by the diagonalization of the eigenvalue problem. Because the eigenproblem is a square matrix, the eigenproblem provides as many eigenmodes and associated eigenvalues as the dimension of matrix $\underline{\mathbf{A}}$. In the optimal case we get $\bar{N} = 2M$ forward modes and the same number of backward modes as already discussed in Sec. 5.2.1.

An alternative to the full matrix diagonalization using LAPACK is the solution for a small finite number of eigenmodes only. This is useful if one is not interested in scattering solutions but, for example, in the determination of the guided eigenmodes of a waveguide structure. A library that provides this functionality is ARPACK (ARnoldi PACKage) [83]. The algorithm is based on the Arnoldi process called the Implicitly Restarted Arnoldi Method. It is most appropriate for large sparse matrices. Unfortunately, the system matrix operators of the FMM are not sparse but dense. Therefore, the performance gain is not very large. The efficiency with respect to the number of calculated solutions is actually quite low. But if only a few eigenmodes are needed, the speedup is roughly in the order of a factor of two to three.

With the eigenmode matrices and the eigenvalues at hand for each layer, we can next apply the scattering matrix algorithm to determine the unknown amplitudes $\mathbf{u}$ and $\mathbf{d}$.

6.4. Scattering Matrix

In the Fourier modal method the matching of the transverse fields at layer interfaces is done using the matching by basis functions introduced in Sec. 5.3.1. Since the eigenmode matrices $\underline{\tilde{\mathbf{M}}}^{(l)}$ are square, the matrix inversion can be carried out using standard techniques provided by LAPACK — we use routines *zgetrf* and *zgetri* [82], or the respective MKL implementation [42] on Intel machines.

The advantages of this matching method is that the fields match exactly everywhere on the planar interface. Empirical studies done for the B-spline modal method (BMM) show that the S-matrix algorithm with the basis function matching usually converges better[2] than the pointwise matching [53, 74]. However, it requires the same basis functions in each layer which also implies the same coordinate meshes. In plain FMM the mesh is an equidistant Cartesian mesh necessary for efficient discretization of the real-space unit cell using the FFT. This Cartesian mesh is naturally used in every layer. Hence, the restriction is not important for plain FMM.

The disadvantage of the Cartesian mesh is that any non-grid-aligned structure is staircase approximated (cf. Fig. 7.1(a)). Furthermore, the number of sampling points N_{fft} must be carefully chosen, such that the size of the structure is correctly represented [84].

These obstacles can be overcome with non-Cartesian meshes that are specifically adapted to the structure. This extension to the FMM towards adaptive coordinates and adaptive spatial resolution using coordinate transformations is topic of Chap. 7.

6.5. Field Sources

After the scattering matrix is calculated, the next step in the structure simulation is the specification of the light source and thereby the amplitudes $\mathbf{u}$ and $\mathbf{d}$ in the source layer. We differentiate mainly between three types of light sources: Plane waves incident from homogeneous semi-infinite half-spaces, (guided) eigenmodes of a waveguide-like structure, and light emission from a dipole within the structure.

6.5.1. Plane Waves

By design, the FMM is constructed to calculate stacked grating structures which give rise to multiple diffraction (cf. Sec. 3.5). Hence, the usual application in the field of nanophotonics are structures such as photonic crystals and metamaterials. They are almost perfectly suited to the requirements the FMM poses. These structures are periodic and their lateral extent is usually sufficiently large so that finite size effects are negligible and can be very well approximated by infinite periodicity. In the propagation direction the structures are finite, as required, and often enclosed by layers of homogeneous superstrate and substrate materials, e.g., air and silicon wavers. The typical light

[2] More precisely, pointwise matching with properly chosen matching points can give comparable convergence rates as the basis function matching, but is is not easy to choose the "correct" matching points.

source for such setups is a plane wave with wave vector $\mathbf{k} = (\alpha_0, \beta_0, \gamma_0)^T$ incident onto the structure from one side under azimuthal angle Θ to the surface normal and polar angle Φ to the x^1–axis. The corresponding field expansion in the homogeneous half-spaces is given by the Rayleigh expansion[3] introduced in Eqs. (3.46), with components of the incident wave vector given by

$$\alpha_0 = k\,\cos\big(\Phi\big)\sin\big(\Theta\big)\ , \tag{6.15a}$$

$$\beta_0 = k\,\sin\big(\Phi\big)\sin\big(\Theta\big)\ , \tag{6.15b}$$

$$\gamma_0 = k\,\cos\big(\Theta\big)\ , \tag{6.15c}$$

where k is the wavenumber in the incoming half-space.

We distinguish between the polarization states of the incident plane wave, s-pol and p-pol (cf. Sec. 2.4.3), which have in general different transmittance and reflectance spectra[4]. In the numerical framework, however, we rather use the descriptions TE and TM, which correspond to s-pol and p-pol for oblique incidence ($\Theta \neq 0$), respectively. The incoming field polarizations are defined as

$$\underline{\text{TE:}} \qquad \mathbf{E}_{\text{inc}} = E_0 \begin{pmatrix} \sin\big(\Phi\big) \\ -\cos\big(\Phi\big) \\ 0 \end{pmatrix}\ , \qquad \mathbf{H}_{\text{inc}} = H_0 \begin{pmatrix} \cos\big(\Phi\big)\cos\big(\Theta\big) \\ \sin\big(\Phi\big)\cos\big(\Theta\big) \\ -\sin\big(\Theta\big) \end{pmatrix}\ , \tag{6.16a}$$

$$\underline{\text{TM:}} \qquad \mathbf{E}_{\text{inc}} = E_0 \begin{pmatrix} \cos\big(\Phi\big)\cos\big(\Theta\big) \\ \sin\big(\Phi\big)\cos\big(\Theta\big) \\ -\sin\big(\Theta\big) \end{pmatrix}\ , \qquad \mathbf{H}_{\text{inc}} = H_0 \begin{pmatrix} -\sin\big(\Phi\big) \\ \cos\big(\Phi\big) \\ 0 \end{pmatrix}\ . \tag{6.16b}$$

The reason for this naming is that in the case of normal incidence, $\Theta = 0$, s-pol and p-pol are not defined anymore, but TE and TM are still meaningful with respect to Eq. (6.16), because in TE (TM) polarization the $\mathbf{E}$ ($\mathbf{H}$) field is entirely in the transverse x^1-x^2-plane.

The eigenmode amplitudes are obtained in a simple fashion: We set $\mathbf{d}_L = \mathbf{0}$ and all entries of $\mathbf{u}_1$ equally to zero, except $u_m^{(1)} = 1$ where $m = \{m_1, m_2\} = \{0, 0\}$ is the index corresponding to the zeroth diffraction order.

6.5.2. Guided Eigenmodes

A second light source, often used in waveguide applications, is the excitation of an eigenmode of the structured first layer. The scattering matrix then couples the eigenmodes, and the transmitted and reflected fields are not separated by the Bragg orders of diffraction into homogeneous layers anymore, but by the eigenmodes — guided or not-guided — of the respective input layer (1) and output layer (L). The necessary task is to find index m of the desired mode and set the respective eigenmode amplitude $u_m^{(1)} = 1$. Depending on the system, this can be a difficult task which is discussed in further detail in Chap. 8.

[3] The Rayleigh expansion is the eigenmode expansion for homogeneous layers, since the eigenmodes of those layers are plane waves.

[4] Unless the structure is symmetric.

The periodic boundary conditions inherent to the method, of course, lead to a periodic arrangement of the waveguides, too. Still, guided eigenmodes may often be represented quite well, if one chooses the lattice spacing sufficiently large, because their fields decay exponentially. However, leaky modes and radiation modes which occur in open systems are replaced by Bloch modes in the eigenmode spectrum. Furthermore, as soon as scattering comes into play, there is undesired crosstalk between the modes, and the results are contaminated by lattice effects. To avoid these shortcomings we introduce open boundary conditions as an extension for the FMM in Sec. 7.2.

6.5.3. Dipoles

Instead of illuminating the structure from the outside, we can also simulate a dipole source inside the structure and calculate the emitted fields in the first and last layer or within the structure. The procedure is extensively discussed in the master's theses of Klock and Lutz [59, 85]. Here, we only give a brief introduction and shortly mention the limits.

The core idea involves two steps. First, the dipole situated at point $\mathbf{r}_0 = (\mathbf{r}_0^{\parallel}, x_0^3)^T$ can be represented by a (free) charge current density

$$\mathbf{j}(\mathbf{r}) = \mathbf{j}_0\, \delta(\mathbf{r}^{\parallel} - \mathbf{r}_0^{\parallel})\, \delta(x^3 - x_0^3) = \mathbf{p}\, \delta(x^3 - x_0^3) \tag{6.17}$$

in the unit cell, with a constant vector $\mathbf{j}_0$ determining strength and orientation of the dipole. Similar to the fields, the current density is expanded into a Floquet-Fourier series with Fourier coefficient given by the two-dimensional lattice Fourier transform of the periodically continued real-space current density. From Eq. (3.18) we obtain

$$\tilde{j}_m^{\rho} = \frac{1}{a_1 a_2} \int_0^{a_1} \mathrm{d}x^1 \int_0^{a_2} \mathrm{d}x^2 \left(\sum_{n=1}^{\infty} j^{\rho}(\mathbf{r} - \mathbf{R}_n) \right) e^{-\mathrm{i}\,\mathbf{G}_m \cdot \mathbf{r}^{\parallel}} = \frac{j_0^{\rho}}{a_1 a_2}\, e^{-\mathrm{i}\mathbf{G}_m \mathbf{r}_0^{\parallel}}\, \delta(x^3 - x_0^3)\,. \tag{6.18}$$

As usual, the coefficients can be represented by a truncated Fourier vector

$$\tilde{\mathbf{j}}^{\rho} = \tilde{\mathbf{p}}^{\rho}\, \delta(x^3 - x_0^3)\,. \tag{6.19}$$

Second, the source is incorporated into the scattering matrix algorithm via a modification of the continuity conditions, Eq. (5.37). This modification can be derived from Maxwell's equations including sources. Here, we skip some steps. For this to work, we must introduce an additional interface at position x_0^3. Let us assume the layers left and right of this new interface are layers (l) and $(l+1)$, respectively. Hence, in truncated reciprocal space the modified continuity condition reads

$$\underline{\mathbf{M}}^{(l)}(x^1, x^2) \begin{pmatrix} \mathbf{u}_-^{(l)} \\ \mathbf{d}_-^{(l)} \end{pmatrix} - \underline{\mathbf{M}}^{(l+1)}(x^1, x^2) \begin{pmatrix} \mathbf{u}_+^{(l+1)} \\ \mathbf{d}_+^{(l+1)} \end{pmatrix} \stackrel{!}{=} \begin{pmatrix} \underline{\boldsymbol{\beta}}\, \underline{\hat{\boldsymbol{\varepsilon}}}^{33}\, \tilde{\mathbf{p}}^3 \\ -\underline{\boldsymbol{\alpha}}\, \underline{\hat{\boldsymbol{\varepsilon}}}^{33}\, \tilde{\mathbf{p}}^3 \\ -\tilde{\mathbf{p}}^2 \\ \tilde{\mathbf{p}}^1 \end{pmatrix}. \tag{6.20}$$

The artificial interface splits the whole structure into two parts described by the S-matrices

$$\begin{pmatrix} \mathbf{u}_l \\ \mathbf{d}_1 \end{pmatrix} = \underline{\mathbf{S}}(1, l) \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{d}_l \end{pmatrix}, \qquad \text{and} \qquad \begin{pmatrix} \mathbf{u}_L \\ \mathbf{d}_{l+1} \end{pmatrix} = \underline{\hat{\boldsymbol{S}}}(l+1, L) \begin{pmatrix} \mathbf{u}_{l+1} \\ \mathbf{d}_L \end{pmatrix}. \tag{6.21}$$

Starting from Eq. (6.20) and assuming $\mathbf{u}_1 = \mathbf{d}_L = \mathbf{0}$, it is possible to derive consistently amplitudes $\mathbf{u}_{l+1}$ and $\mathbf{d}_l$, which incorporate multiple scattering from both structure parts. Taking them as input amplitude for the scattering matrix of the respective subsystem, the two subsystems can be treated independently and can be solved for all desired quantities as described in Sec. 5.3. For instance, the amplitudes of the emitted light in the first and last layer are obtained as $\mathbf{d}_1 = \underline{\mathbf{S}}_{22}\mathbf{d}_l$ and $\mathbf{u}_L = \underline{\hat{\boldsymbol{S}}}_{11}\mathbf{u}_{l+1}$, respectively.

The described charge current density is lattice periodic which means that we actually simulate a dipole in each unit cell which are all radiating coherently. Alternatively, it is possible to simulate only a single source in the whole structure by use of charge current Fourier vectors that are shifted by a lateral $\mathbf{k}^{\parallel}$-vector

$$\tilde{\mathbf{j}}^{\rho}(\mathbf{k}^{\parallel}, x^3) = \tilde{\mathbf{p}}^{\rho}(\mathbf{k}^{\parallel})\,\delta(x^3 - x_0^3)\,. \tag{6.22}$$

The calculation from above must be repeated for a sufficient number of sampling points in the first BZ, and the calculated fields $\mathbf{E}(\mathbf{r}, \mathbf{k}^{\parallel})$ and $\mathbf{H}(\mathbf{r}, \mathbf{k}^{\parallel})$ at every point of interest are finally integrated over $\mathbf{k}^{\parallel}$ in order to obtain the real-space electromagnetic fields of a single source in an infinite periodic structure. This procedure can be interpreted as follows: Due to the modification, the current density remains (Bloch-) periodic, but every lattice site gains an additional phase $e^{\mathrm{i}\mathbf{k}^{\parallel}\mathbf{R}_n}$. The integration over the BZ superposes the emissions such that they destructively interfere and only the emission of a single dipole source remains.

It is necessary to comment on the performance of dipole sources. At the origin of dipole sources, the fields diverge. This is not very nicely representable with a finite Fourier series. Especially in a two-dimensional lattice (three-dimensional problem) the limited number of plane waves that can be used in the simulation is often not high enough to get converged results. Unfortunately, empiric studies indicate that the simulation of the single point source with the additional two-dimensional integration over the first BZ is far beyond the scope of what is reasonably achievable under these circumstances. However, a promising ansatz to increase the convergence performance is the use of adaptive meshes as presented in Chap. 7.

6.6. Transmittance and Reflectance

Transmittance T and reflectance R are defined as

$$T = \frac{P_{\mathrm{trans}}}{P_{\mathrm{inc}}} \qquad \text{and} \qquad R = \frac{P_{\mathrm{refl}}}{P_{\mathrm{inc}}}\,, \tag{6.23}$$

which are the portions of power P carried by the electromagnetic fields which reach the respective sides of the structure traveling forward and backward, respectively, in relation to the power carried by the incident field. Since the structures we consider are infinitely extended in the lateral plane and periodic, it seems reasonable to consider only the part of the transported power which travels orthogonal to the structure's surfaces. The power flux of harmonic solutions is given by the time averaged Poynting vector which has been introduced in Eq. (2.23) and for curvilinear coordinate systems in Eq. (2.91). According to our considerations above, we are interested in its x^3-component (z-component) integrated over the unit cell

$$P = \int_{UC} \mathrm{d}A\; S^3(\mathbf{r}, \omega) = \frac{1}{2}\int_{UC} \mathrm{d}A\; \mathrm{Re}\left(\left(\mathbf{E}(\mathbf{r}, \omega) \times \mathbf{H}^*(\mathbf{r}, \omega)\right) \cdot \hat{\mathbf{x}}^3\right)\,. \tag{6.24}$$

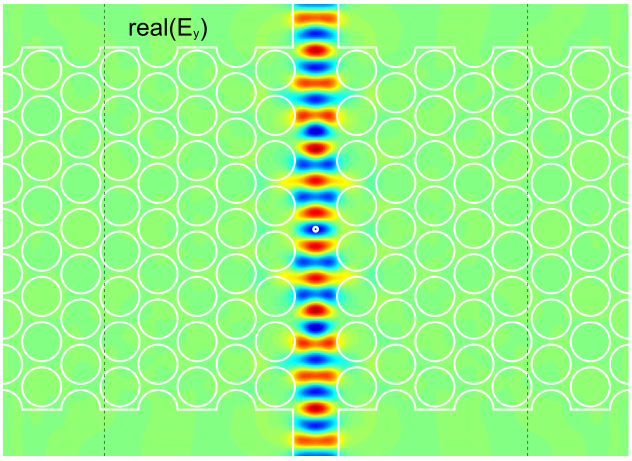


Figure 6.1.: *Dipole source emission pattern calculated with the FMM. Out of plane (y-) oriented dipole within a PC slab waveguide (white mark). The emission frequency is chosen within the bandgap of the PC. The unit cell is isolated with absorbing boundaries (cf. Sec. 7.2). Picture from Ref. [59].*

This relation is very familiar from the mode orthogonality conditions, Eqs. (4.26), in the case of the reciprocity theorem with conjugated fields, Eq. (4.15). We recall that the orthogonality conditions with the fields' conjugated form are only valid for systems with real material parameters. We can make use of the orthogonality relations for such systems by considering the fields' eigenmode expansions (cf. Eq. (5.17) and Eq. (5.18), or Eq. (5.20)). For the illustration, we pick the transmitted fields which are given by the forward terms in layer L with amplitudes $u_m^{(L)}$

$$\begin{aligned} P_{\text{trans}}(\omega) &= \frac{1}{2}\int_{UC} \mathrm{d}A\ \mathrm{Re}\left(\left(\sum_j u_j^{(L)}\, \mathbf{E}_j^{(L)}(\mathbf{r},\omega) \times \sum_k u_k^{*(L)}\, \mathbf{H}_k^{*(L)}(\mathbf{r},\omega) \right) \cdot \hat{\mathbf{x}}^3 \right) \\ &= \frac{1}{2}\mathrm{Re}\left(\sum_{j,k} u_j^{(L)} u_k^{*(L)} \underbrace{\int_{UC} \mathrm{d}A\ \left(\mathbf{E}_j^{(L)}(\mathbf{r},\omega) \times \mathbf{H}_k^{*(L)}(\mathbf{r},\omega) \right) \cdot \hat{\mathbf{x}}^3}_{\delta_{jk}} \right) \\ &= \frac{1}{2}\sum_j |u_j^{(L)}|^2 = \sum_j P_{\text{trans},j}\,, \end{aligned} \tag{6.25a}$$

where we assumed normalized eigenmodes. In the same way we can calculate the reflected power

$$P_{\text{refl}}(\omega) = \frac{1}{2}\sum_j |d_j^{(1)}|^2 = \sum_j P_{\text{refl},j}\,, \tag{6.25b}$$

and the incoming power

$$P_{\text{inc}}(\omega) = \frac{1}{2}\sum_j |u_j^{(1)}|^2\,, \tag{6.25c}$$

per unit cell in the first layer. This definition is valid for all kind of eigenmodes in the first and last layer, no matter whether they are plane wave solutions of the Rayleigh expansion in homogeneous half-spaces, or eigenmodes of structured input and output regions, as long as the layers consist of materials with real parameters. As a consequence, we can also define a transmittance and reflectance into the j-th eigenmode by

$$T_j = \frac{P_{\text{trans},j}}{P_{\text{inc}}} \qquad \text{and} \qquad R_j = \frac{P_{\text{refl},j}}{P_{\text{inc}}}\,. \tag{6.26}$$

The sum over the transmittance and reflectance into the distinct eigenmodes must be $\sum_j T_j + R_j = 1$ because of energy conservation if the structure is non-absorbing as well. In case of homogeneous input and output regions the j-th mode corresponds to the m-th Bragg order. Please note that the derivation was carried out independent of the used coordinate system.

However, if the layers involve complex material parameters, the mode orthonormality condition used in Eq. (6.25a) is not applicable anymore. Hence, Eq. (6.24) leads for the transmitted power to

$$\begin{aligned} P_{\text{trans}}(\omega) &= \frac{1}{2}\text{Re}\left(\sum_{j,k} u_j u_k^* \int\limits_{UC} \text{d}A \, \left(\mathbf{E}_j(\mathbf{r},\omega) \times \mathbf{H}_k^*(\mathbf{r},\omega)\right) \cdot \hat{\mathbf{x}}^3 \right) \\ &= \frac{1}{2}\text{Re}\left(\sum_{j,k} u_j u_k^* \int\limits_{UC} \text{d}A \, \frac{1}{\sqrt{g(\mathbf{r})}} \left(E_{1,j} H_{2,k}^* - E_{2,j} H_{1,k}^*\right)\right). \end{aligned} \tag{6.27}$$

Here, from first to second line, we used the definition of the covariant cross product (cf. Eq. (2.68) and Eq. (2.80)). Next, we use that in case of $x^3 = \bar{x}^3 = z$ we get $\frac{\text{d}A}{\sqrt{g}} = \text{d}x^1\text{d}x^2$. Again, we conveniently leave out the layer labels for brevity. Substituting the plane wave expansion of the forward traveling eigenmodes in layer L

$$E_{\rho,j}^+(x^1,x^2,x^3) = \sum_m \tilde{E}_{\rho,m,j} \; e^{+\text{i}\alpha_{m_1} x^1 + \text{i}\beta_{m_2} x^2} \; e^{+\text{i}\gamma_j (x^3 - x_{L-1}^3)}\,, \tag{6.28a}$$

$$H_{\rho,k}^{*+}(x^1,x^2,x^3) = \sum_n \tilde{H}_{\rho,m,k}^* \; e^{-\text{i}\alpha_{n_1}^* x^1 - \text{i}\beta_{n_2} x^2} \; e^{-\text{i}\gamma_k^* (x^3 - x_{L-1}^3)}\,, \tag{6.28b}$$

into Eq. (6.27), and using the orthonormality relation Eq. (6.2) in order to get rid of the integral over the exponential terms, we end up with

$$\begin{aligned} P_{\text{trans}}(\omega) &= \frac{1}{2}\text{Re}\left(\sum_{j,k} u_j u_k^* \underbrace{\sum_m \left(E_{1,m,j} H_{2,m,k}^* - E_{2,m,j} H_{1,m,k}^*\right)}_{\mathsf{S}_{jk}} \; e^{\text{i}(\gamma_j - \gamma_k^*)x^3}\right) \\ &= \sum_j \underbrace{\frac{1}{2}|u_j|^2 \; \text{Re}(\mathsf{S}_{jj})}_{\substack{\text{Transmittance into} \\ j\text{-th mode } P_{\text{trans},j}}} e^{-2\text{Im}(\gamma_j)x^3} \; + \; \underbrace{\frac{1}{2}\text{Re}\left(\sum_j \sum_{k\neq j} u_j u_k^* \; \mathsf{S}_{jk} \; e^{\text{i}(\gamma_j - \gamma_k^*)x^3}\right)}_{\text{Interference}}. \end{aligned} \tag{6.29}$$

Similar expressions are obtained for the reflected and incident fields. For the guided modes we can define a transmittance and reflectance as in Eq. (6.26). Please note that, different to the case of purely non-absorbing materials above, the obtained expression contains terms which describe the

interference between eigenmodes. These interference terms can provide negative contributions to the overall power. Hence, the sum over transmittance and reflectance into the modes $\sum_j T_j + R_j \neq 1$ and can even be larger than 1. This means that transmittance (reflectance) into the modes does not have the same physical meaning as the total transmittance (reflectance). Nevertheless, these quantities are useful in many cases, since the deviations from the non-absorbing case with perfect orthonormality of the modes is rather small. We would like to remark that as soon as absorbing materials are involved, the total transmittance and reflectance $T + R < 1$ do not sum up to one anymore.

6.7. Field Reconstruction

The reconstruction of the fields is a task that must be performed for each layer l of the system separately. If the amplitudes $\mathbf{u}^{(l)}$ and $\mathbf{d}^{(l)}$ have been obtained from the scattering matrix (cf. Sec. 5.3.7), the field Fourier vector $\tilde{\mathbf{V}}$ is given by the analog of Eq. (5.22) in Fourier space

$$\tilde{\mathbf{V}}^{(l)}(x^3) = \underline{\tilde{\mathbf{M}}}^{(l)} \underbrace{\underline{\mathbf{\Phi}}^{(l)}(x^3) \begin{pmatrix} \mathbf{u}^{(l)} \\ \mathbf{d}^{(l)} \end{pmatrix}}_{\text{rephased amplitudes}} = \left(-\tilde{\mathbf{E}}_2^{(l)}(x^3),\, \tilde{\mathbf{E}}_1^{(l)}(x^3),\, \tilde{\mathbf{H}}_1^{(l)}(x^3),\, \tilde{\mathbf{H}}_2^{(l)}(x^3) \right)^T \tag{6.30}$$

for both the large and small eigenproblem, where vector $\tilde{\mathbf{V}}^{(l)}$ contains *all* $4M$ phased Fourier coefficients of the four transverse field components at coordinate x^3.

The corresponding longitudinal components are calculated from the discretized version of Eqs. (5.5)

$$\tilde{\mathbf{H}}_3^{(l)}(x^3) = \underline{\mathbf{U}}_{\mathsf{H}}^{(l)}\, \tilde{\mathbf{V}}^{(l)}(x^3) \qquad \text{with } \underline{\mathbf{U}}_{\mathsf{H}}^{(l)} = \left(-\frac{1}{\omega^2} \underline{\hat{\boldsymbol{\mu}}}^{33}\, \underline{\boldsymbol{\alpha}}\,, -\frac{1}{\omega^2} \underline{\hat{\boldsymbol{\mu}}}^{33}\, \underline{\boldsymbol{\beta}}\,, -\underline{\hat{\boldsymbol{\mu}}}^{31}\,, -\underline{\hat{\boldsymbol{\mu}}}^{32} \right), \tag{6.31a}$$

$$\tilde{\mathbf{E}}_3^{(l)}(x^3) = \underline{\mathbf{U}}_{\mathsf{E}}^{(l)}\, \tilde{\mathbf{V}}^{(l)}(x^3) \qquad \text{with } \underline{\mathbf{U}}_{\mathsf{E}}^{(l)} = \left(\underline{\hat{\boldsymbol{\varepsilon}}}^{32}\,, -\underline{\hat{\boldsymbol{\varepsilon}}}^{31}\,, +\underline{\hat{\boldsymbol{\varepsilon}}}^{33}\, \underline{\boldsymbol{\beta}}\,, -\underline{\hat{\boldsymbol{\varepsilon}}}^{33}\, \underline{\boldsymbol{\alpha}} \right). \tag{6.31b}$$

In the small eigenproblem case matrices $\underline{\hat{\boldsymbol{\varepsilon}}}^{31}$, $\underline{\hat{\boldsymbol{\varepsilon}}}^{32}$, $\underline{\hat{\boldsymbol{\mu}}}^{31}$, and $\underline{\hat{\boldsymbol{\mu}}}^{32}$ vanish. Hence, with Eqs. (6.30) and Eqs. (6.31), the Fourier coefficients of all six field components are at hand. There are several ways to calculate the real-space electromagnetic fields from these coefficients. They are introduced in the following.

6.7.1. Fourier Series

The evident way for the real-space field reconstruction is the calculation of the truncated Fourier series, Eq. (6.4), for specific lateral coordinates $\mathbf{r}^{\parallel} = (x^1, x^2)$. This is the way of choice as long as the number of requested lateral coordinates or the number of plane waves is small, because the method is easy but cumbersome.

6.7.2. Inverse Fourier Transform

The fastest way to get the field distribution in the whole unit cell or for a large number of coordinates is the inverse Fast Fourier Transform (iFFT), which is similar to a FFT on the Fourier coefficients, but

with a positive exponential and without the factor $1/N_{\mathrm{fft}}$ in the underlying equation (cf. Eq. (3.22)). As the FFT, the iFFT we use is provided by the open source C subroutine library FFTW [40, 41], and on Intel machines by the routines of the Intel Math Kernel Library [42]. They are available for transformations in one and two dimensions.

The drawback of this method is that the fields are only calculated on a regular grid with equidistant spacing Δ_1 and Δ_2 in both dimensions. The spacing can be chosen by the number of sampling points N_{fft} in the respective direction (cf. Eqs. (3.21)). The transformation is on a matrix of the Fourier coefficients $\tilde{f}_m$ ordered in a particular way. Matrix entries with no corresponding coefficients are set to zero. Lake for all truncated Fourier series, the obtained field distributions feature small spatial oscillations with a wavelength corresponding to the highest non-zero Fourier order (similar as for the reconstructed permittivity which is illustrated in Fig. 6.2(b)). Noticable are in particular field overshoots at material discontinuities, where field components are either discontinuous as well or have at least discontinuous dervatives. This ringing is an expression of Gibbs's phenomenon (cf. Sec. 3.4.3).

6.7.3. Singular Fourier Pade

A rather innovative approach for the field reconstruction is the application of a *singular Fourier-Padé* (SFP) approximation as described by Driscoll and Fornberg [86].

Discontinuities in the field distributions or their derivatives — as they occur at material interfaces — cause Gibbs' phenomenon. Consequently, the truncated Fourier series fails to converge at such jump positions. The core idea of the SFP approximation is to use information about the jump locations to increase the convergence rate. Instead of calculating the Fourier series, the method uses the Fourier coefficients to derive a polynomial approximation to the exact function. This process is essentially broken down to the solution of a linear system for the polynomials' coefficients.

The truncated one-dimensional Fourier series (cf. Eq. (3.17)) can be rewritten into Laurent expansions

$$f(x) = \sum_{k=-M_p}^{+M_p} \tilde{f}_k \, e^{\mathrm{i}k\frac{2\pi}{a}x} \quad \xrightarrow{z=e^{\mathrm{i}\frac{2\pi}{a}x}} \quad f(z) = \sum_{k=0}^{M_p} \tilde{f}_k \, z^k + \sum_{k=0}^{M_p} \tilde{f}_{-k} \, z^{-k} = f^+(z) + f^-(z^{-1}) \,, \tag{6.32}$$

where the zeroth coefficient is halfed and distributed between the two sums. The truncated Taylor polynomials f^+ and f^- are replaced by Padé polynomials $p^+(z)$, $q^+(z)$, $p^-(z)$, $q^-(z)$, each of degree $N_p/2$, such that

$$p^{\pm}(z) - q^{\pm}(z) f^{\pm}(z) = \mathcal{O}(z^{N_p+1}) \quad \text{for } z \to 0 \,. \tag{6.33}$$

Polynomials that fulfill Eq. (6.33) give rise to the Fourier-Padé approximation of function f, which is given by

$$f(x) \approx \frac{p^+(e^{\mathrm{i}\frac{2\pi}{a}x})}{q^+(e^{\mathrm{i}\frac{2\pi}{a}x})} + \frac{p^-(e^{-\mathrm{i}\frac{2\pi}{a}x})}{q^-(e^{-\mathrm{i}\frac{2\pi}{a}x})} \,. \tag{6.34}$$

Driscoll and Fornberg show that jump discontinuities in $f(x)$ lead to logarithmic singularities in $f^{\pm}$ which are not well representable with Padé polynomials. Therefore, these logarithmic singularities

are *explicitly* incorporated into the Fourier-Padé approximant, Eq. (6.34). If jump discontinuities occur at positions $z_1, \ldots, z_s$, the new SFP approximant is given by

$$f(x) = f^+\Big(z(x)\Big) + f^-\Big(z^{-1}(x)\Big)\,, \tag{6.35a}$$

with

$$f^\pm(z) = \frac{p^\pm(z)}{q^\pm(z)} + \frac{r_1^\pm(z)}{q^\pm(z)} \log\left(1 - \frac{z}{z_1^{\pm 1}}\right) + \ldots + \frac{r_s^\pm(z)}{q^\pm(z)} \log\left(1 - \frac{z}{z_s^{\pm 1}}\right) + \mathcal{O}(z^{N_p+1})\,. \tag{6.35b}$$

In order to make the coefficients well-determined, the orders of the polynomials denoted by n_p, n_q, and n_{r_i} with $i = 1, \ldots, s$ ($n_r = \sum_{i=1}^{s} n_{r_i}$), respectively, must sum up to $N_p - 1$.

For the one-dimensional SFP there exists the MATLAB function *padelog* written by Driscoll [87] which calculates the polynomial coefficients from the Fourier coefficients and the coordinates of the jump locations. Since we need two-dimensional field reconstructions, we worked on the application of the procedure to two-dimensional problems. Our ansatz is the successive application of one-dimensional reconstructions in x^1 and x^2 directions. The exemplary studies depicted in Fig. 6.2 show the improved permittivity reconstruction using SFP in comparison to iFFT for a square waveguide structure. These results demonstrate the significant potential of the SFP reconstruction.

However, a crucial issue is the automatic detection of the jump locations, considering that the SFP is designed for exact Fourier coefficients, but we have only approximated aliased Fourier coefficients available on a discretized grid. Hence, the numerical jump coordinates can only be determined with a deviation in the order of the sampling interval Δ (cf. Eq. (3.21b)). We noticed in our studies that the error of the reconstructed functions is very sensitive to small deviations between the jump location encoded in the truncated Fourier series and the automatically detected positions. This sometimes leads to artifacts at material interfaces whose magnitude can become comparable to the deviations of the truncated Fourier series. A different number of Fourier coefficients can already improve the situation. Still, such deviations occur only at misplaced jumps. Everywhere else, the SFP converges much faster and lacks the oscillations inherent to Fourier series and iFFT.

Due to these obstacles, we could not establish a reliable field reconstruction routine for our numerical framework in the course of this work. This is a task which remains for future development.

6.8. Symmetry Reduction

As discussed in Chap. 5, the basic principle of modal methods is to expand the electromagnetic fields $\mathbf{\Psi}$ of a layer into its eigenmodes $\mathbf{\Psi}_j$ (cf. Eq. (5.1)). These eigenmodes are solutions of Maxwell's equations for a given frequency ω and the layer's permittivity $\underline{\varepsilon}(\mathbf{r})$ and permeability $\underline{\boldsymbol{\mu}}(\mathbf{r})$ defining the layer's structure. If this structure is invariant under the application of a certain set of symmetry operations — like rotations or mirror reflections — it suggests itself that the eigenmodes $\mathbf{\Psi}_j$ of this structure inherit these symmetry properties as well.

The idea of *symmetry reduction* is to use the symmetry properties of the investigated layer to reduce the number of unknown Fourier coefficients N_{ep} in the discretized eigenvalue problems,[5] Eq. (6.11)

[5] Large eigenproblem: $N_{\mathrm{ep}} = 4M$, small eigenproblem: $N_{\mathrm{ep}} = 2M$.

and Eqs. (6.13). Since any (dense) eigenproblem asymptotically scales with $\mathcal{O}(N_{\text{ep}}^3)$ with respect to the number of operations necessary for its diagonalization, a significant reduction of unknowns leads to a considerable improvement in the computational performance. The obtained results are fully equivalent to the solutions of the full problem, because only redundant information is cropped.

In the FMM, the symmetry of a (periodic) structure splits into two parts: the symmetry of the lattice and the symmetry of the structure pattern within the unit cell. However, due to the numerical discretization, there is a third aspect that needs to be considered: the symmetry of the mesh (cf.

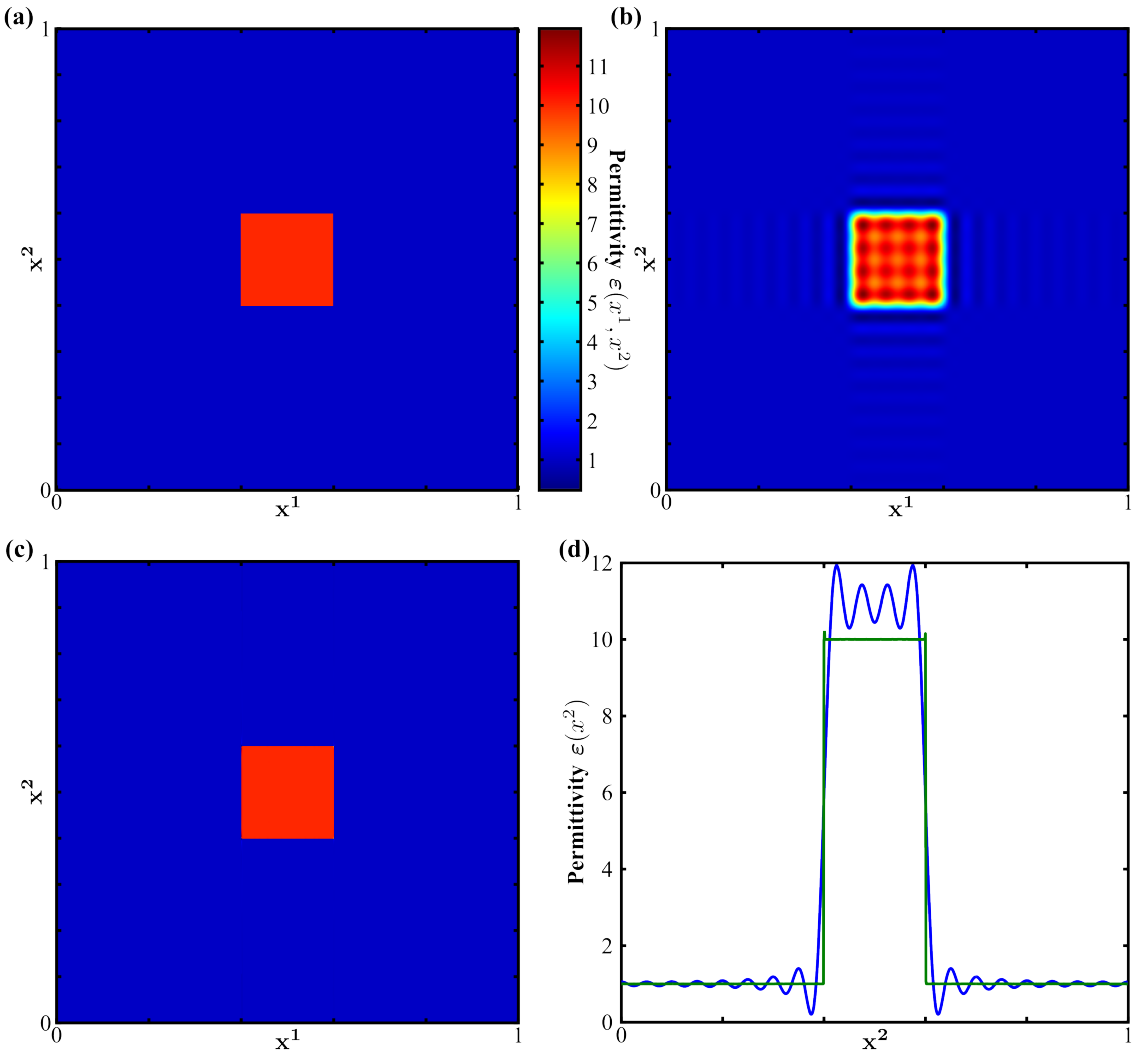


Figure 6.2.: *Reconstruction of a square permittivity distribution within the unit cell (color coded). The Fourier coefficients were obtained with a FFT with* $N_{\text{fft}} = 1024$ *sampling points per direction. The plots have* 1000 *points per dimension. (a) Original function* $\varepsilon(x^1, x^2)$*. (b) Reconstruction by an inverse FFT with* $41 \times 41 = 1681$ *retained Fourier coefficients. (c) Reconstruction with SFP from the same Fourier coefficients as in (b). Panels (a)-(c) share the same color scale. (d) Comparison between iFFT (blue) and SFP (green) along the cut* $x^1 = 0.426$*.*

Plane point group	Symmetry operations
C_1	$\hat{e}$
C_s	$\hat{e}$, $\hat{\sigma}_v$
C_2	$\hat{e}$, $\hat{c}_2$
C_{2v}	$\hat{e}$, $\hat{c}_2$, $\hat{\sigma}_v$, $\hat{\sigma}_d$
C_3	$\hat{e}$, $\hat{c}_3^1$, $\hat{c}_3^2$
C_{3v}	$\hat{e}$, $\hat{c}_3^1$, $\hat{c}_3^2$, $\hat{\sigma}_b$, $\hat{\sigma}_d$, $\hat{\sigma}_f$
C_4	$\hat{e}$, $\hat{c}_4^1$, $\hat{c}_4^3$, $\hat{c}_2$
C_{4v}	$\hat{e}$, $\hat{c}_4^1$, $\hat{c}_4^3$, $\hat{c}_2$, $\hat{\sigma}_{vx}$, $\hat{\sigma}_{vy}$, $\hat{\sigma}_{dx}$, $\hat{\sigma}_{dy}$
C_6	$\hat{e}$, $\hat{c}_6^1$, $\hat{c}_6^5$, $\hat{c}_3^1$, $\hat{c}_3^2$, $\hat{c}_2$
C_{6v}	$\hat{e}$, $\hat{c}_6^1$, $\hat{c}_6^5$, $\hat{c}_3^1$, $\hat{c}_3^2$, $\hat{c}_2$, $\hat{\sigma}_{vx}$, $\hat{\sigma}_{vA}$, $\hat{\sigma}_{vB}$, $\hat{\sigma}_{dy}$, $\hat{\sigma}_{dC}$, $\hat{\sigma}_{dD}$

Table 6.1.: *Plane point groups and their symmetry operations [88].*

Sec. 3.4.2). Hence, the symmetry of the problem is the common symmetry of lattice, structure pattern, and mesh. As we will remark later on, there is a fourth dependence: the k-vector of the incident field, which can influence the symmetry of the k-space. However, in this work we only consider normal incidence, such that the symmetry of real-space lattice and reciprocal lattice accord.

In practice, the symmetry properties of the layer are mostly determined by the symmetry of the structure pattern in the unit cell. This is because lattice and mesh in the standard FMM are usually closely connected. The former is often a tetragonal or orthorhombic lattice where the lattice vectors are mutually orthogonal, and the latter is a simple Cartesian grid.[6] A considerable deviation from this close connection possibly comes into play when curvilinear coordinate systems and coordinate transformations are used as will be the case in Chap. 7.

Bai and Li proposed an elegant way to approach symmetry reduction in the FMM by group theoretic considerations [88]. In Sec. 6.8.1, we will only briefly introduce the basics in order to give a foundation for the application of the symmetry reduction to FMM problems. For a comprehensive but still slender introduction into the general topic of group theory we recommend Ref. [89]. Here, we restrict the treatment to a detailed exemplary incorporation of the C_{2v}-symmetry into the FMM.

6.8.1. Basics

We assume the structure of the layer to have the symmetry in the x^1-x^2-plane described by the *plane point group* $G\{g_1, g_2, \ldots, g_{N_s}\}$ with N_s symmetry operators g_n. Table 6.1 gives an overview of plane point groups and their symmetry operators. The first group element $g_1 = \hat{e}$ is the identity. The operators $\hat{c}_p$ describe a rotation with rotation axis along x^3 (normal to the plane of periodicity) and

[6] Historically, lattices with non-orthogonal in-plane lattice vectors, e.g., monoclinic or hexagonal, were treated separately with skew regular equidistant meshes. We did not include the corresponding modified eigenproblems in our considerations, because these problems are nowadays easily treated with the coordinate transformation techniques which will be introduced in Chap. 7. Then, the problem reduces to the mentioned orthogonal lattices in the transformed space with unit cell discretizations on the Cartesian mesh.

rotation angle $\frac{2\pi}{p}$. Mirror reflection operations are denoted by $\hat{\sigma}$, where the subscript label indicates the mirror plane which always includes the x^3 axis.

The matrix representation of symmetry group G in three-dimensional real-space $\mathbb{R}^3$ is given by the set of 3×3 matrices $\{\underline{\mathbf{M}}(g_n) = \underline{\mathbf{M}}_n : n = 1, \ldots, N_s\}$ corresponding to the linear operators $\hat{\mathrm{D}}(g_n)$ of group elements g_n. These matrices transform position vectors $\mathbf{r}$ into symmetrical vectors

$$\mathbf{r}' = \hat{\mathrm{D}}(g_n)\,\mathbf{r} = \underline{\mathbf{M}}_n \mathbf{r}\,. \tag{6.36}$$

Any (electromagnetic) vector field $\mathbf{\Psi}(\mathbf{r})$ complies with the symmetry of group G if for all symmetry operations $g_n \in G$ the relation

$$\hat{\mathrm{D}}(g_n)\,\mathbf{\Psi}(\mathbf{r}) = \underline{\mathbf{M}}_n\ \mathbf{\Psi}\left(\underline{\mathbf{M}}_n^{-1}\mathbf{r}\right)\ \stackrel{!}{=}\ \mathbf{\Psi}(\mathbf{r})\,, \tag{6.37a}$$

and all pseudo vector fields $\mathbf{\Psi}'(\mathbf{r})$ the similar relation

$$\hat{\mathrm{D}}(g_n)\,\mathbf{\Psi}'(\mathbf{r}) = \det(\underline{\mathbf{M}}_n)\ \underline{\mathbf{M}}_n\ \mathbf{\Psi}'\left(\underline{\mathbf{M}}_n^{-1}\mathbf{r}\right)\ \stackrel{!}{=}\ \mathbf{\Psi}'(\mathbf{r}) \tag{6.37b}$$

holds true, where $\underline{\mathbf{M}}_n^{-1}$ denotes the inverse operation. The equivalent relation for tensor quantities is

$$\hat{\mathrm{D}}(g_n)\,\underline{\boldsymbol{\varepsilon}}(\mathbf{r}) = \underline{\mathbf{M}}_n^T\ \underline{\boldsymbol{\varepsilon}}\left(\underline{\mathbf{M}}_n^{-1}\mathbf{r}\right)\ \underline{\mathbf{M}}_n\ \stackrel{!}{=}\ \underline{\boldsymbol{\varepsilon}}(\mathbf{r})\,, \tag{6.38}$$

here at the example of the permittivity tensor.

Representations of the symmetry group are closely connected to a basis of the vector space they are representing in. In the above example the vector space is the three-dimensional Euclidean space $\mathbb{R}^3$ described by the basis vectors $\mathbf{e}_\rho$ of the chosen coordinate system.

Similar as in Euclidean space, the representation of the symmetry group in a field vector space is given by a set of $N_f \times N_f$ matrix operators $\underline{\mathbf{T}}(g_n) = \underline{\mathbf{T}}_n$. The linear dependence of the basis functions $\mathbf{\Psi}^i$ under the symmetry operation,

$$\hat{\mathrm{D}}(g_n)\mathbf{\Psi}^i(\mathbf{r}) = \sum_{j=1}^{N_f} T_{ji}(g_n)\,\mathbf{\Psi}^j(\mathbf{r})\,, \tag{6.39}$$

define the matrix elements of $\underline{\mathbf{T}}_n$.

So far we picked an arbitrary basis for this vector space. However, if it is possible by similarity transformations[7] to obtain a block diagonal form of $\underline{\mathbf{T}}_n$, it is called reducible. If the representations are not further reducible, and all of them have the same form

$$\underline{\mathbf{T}}_n = \begin{bmatrix} \underline{\mathbf{T}}_n^{(1)} & & & \\ & \underline{\mathbf{T}}_n^{(2)} & & \\ & & \ddots & \\ & & & \underline{\mathbf{T}}_n^{(s)} \end{bmatrix} = \bigoplus_{k=1}^{s} \underline{\mathbf{T}}_n^{(k)}\,, \tag{6.40}$$

[7] A similarity transformation is defined as $\underline{\mathbf{T}}'_n = \underline{\mathbf{B}}\,\underline{\mathbf{T}}_n\,\underline{\mathbf{B}}^{-1}$ with non-singular square matrix $\underline{\mathbf{B}}$.

they are called *irreducible*. The corresponding basis is the *canonical basis*. As can be seen in Eq. (6.40), $\underline{\mathbf{T}}_n$ can then be written as direct sum of invariant subspaces $\mathbb{L}^{d_k}$ of dimensionality d_k represented by the unitary block matrices $\underline{\mathbf{T}}_n^{(s)}$. These block matrices are called *irreducible representations* of group G. As a consequence, the coupling between the basis states is weakened to the lowest possible degree in the canonical basis.

Combining Eq. (6.37) and Eq. (6.39), we derive a fundamental equation for the canonical basis vectors $\mathbf{\Psi}^{[i]}$:

$$\underline{\mathbf{M}}_n \; \mathbf{\Psi}^{[i]}\left(\underline{\mathbf{M}}_n^{-1}\mathbf{r}\right) = \sum_{j=1}^{N_f} \left(\underline{\mathbf{T}}_n\right)_{ji} \mathbf{\Psi}^{[j]}(\mathbf{r})\,. \tag{6.41}$$

The block diagonal form of $\underline{\mathbf{T}}_n$ in the canonical basis greatly simplifies this relation. It can be shown that the irreducible representations of the (plane) point groups are always of dimensionality $1 \leq d_k \leq 2$.

In order to obtain the eigenmodes of a certain symmetry, we evaluate Eq. (6.41) and deduce relations for the field's Fourier coefficients. These relations are used to modify the eigenproblem and especially reduce its size.

For the evaluation of Eq. (6.41) and the deduction of symmetry relations for the Fourier coefficients we need the irreducible representations of the symmetry group. Fortunately, they are unambiguously determined by the corresponding *character table* which can be found in many handbooks of group theory, e.g., Ref. [90]. The character tables list the *characters* χ of all irreducible representations. The characters of the matrices are given by their traces

$$\chi^{(k)}(g_n) = \mathrm{Tr}(\underline{\mathbf{T}}_n^{(k)}) = \sum_i \left(\underline{\mathbf{T}}_n^{(k)}\right)_{ii}, \tag{6.42}$$

which are invariant under similarity transformations and, therefore, characteristic for the representation in the linear vector space.

6.8.2. $\mathrm{C_{2v}}$ Symmetry in Detail

We choose the C_{2v}-symmetry [91] for a detailed description of the procedure to obtain the symmetry reduced eigenproblem, because we can apply it to our investigated waveguide systems in Chap. 8. The scheme for the reduction of computational effort for other symmetries is similar to the C_{2v} case. They have been discussed in literature, i.e., in Refs. [92–95].

The C_{2v} symmetry group consists of the symmetry operations $\{\hat{e}, \hat{c}_2, \hat{\sigma}_v, \hat{\sigma}_d\}$ (cf. Tab. 6.1). The matrix representations of the group members $\underline{\mathbf{M}}_n$ in three-dimensional position space with a Cartesian coordinate system are deduced from Eq. (6.36).

We start with the identity operation $\hat{e}$, which is easily written down as the 3×3 identity matrix

$$\underline{\mathbf{M}}_1 = \underline{\mathbf{M}}(\hat{e}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{6.43a}$$

and maps the coordinates onto themselves. Next, the $\hat{c}_2$ rotation operation maps coordinates $x^1 \to -x^1$ and $x^2 \to -x^2$. Thus, the appropriate matrix representation is

$$\underline{\mathbf{M}}_2 = \underline{\mathbf{M}}(\hat{c}_2) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} . \tag{6.43b}$$

A particularity comes into play when we consider reflection operations $\hat{\sigma}_v = \hat{\sigma}_x$ at the axis $x = x^1$, and $\hat{\sigma}_d = \hat{\sigma}_y$ at the axis $y = x^2$. The reflections of a vector (i.e., $\mathbf{r}$ and $\mathbf{E}$) and a pseudo-vector (i.e.,$\mathbf{H}$) must be distinguished. The $\hat{\sigma}_x$ operation changes component $x^2 \to -x^2$ of a vector. The reflection of a pseudo-vector leads to an additional over-all sign flip due to the determinate in Eq. (6.37b). Hence, the matrix representations of $\hat{\mathrm{D}}(\hat{\sigma}_x)$ are

$$\underline{\mathbf{M}}_3 = \begin{cases} \underline{\mathbf{M}}^{nv}(\hat{\sigma}_x) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} & \text{for } \mathbf{r} \text{ and } \mathbf{E}, \text{ and} \\ \underline{\mathbf{M}}^{pv}(\hat{\sigma}_x) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} & \text{for } \mathbf{H}. \end{cases} \tag{6.43c}$$

Similarly, the $\hat{\sigma}_y$ operation changes component $x^1 \to -x^1$ of a vector, and its matrix representations read

$$\underline{\mathbf{M}}_4 = \begin{cases} \underline{\mathbf{M}}^{nv}(\hat{\sigma}_y) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} & \text{for } \mathbf{r} \text{ and } \mathbf{E}, \text{ and} \\ \underline{\mathbf{M}}^{pv}(\hat{\sigma}_y) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} & \text{for } \mathbf{H}. \end{cases} \tag{6.43d}$$

Thus, we obtained the necessary real-space representations of symmetry C_{2v}. Notice that all matrices are self-inverse, i.e., $\underline{\mathbf{M}}_n = \underline{\mathbf{M}}_n^{-1}$.

What remains to be defined in order to solve Eq. (6.41) are the irreducible representations $\underline{\mathbf{T}}_n^{(k)}$ of the group. To this end, we have a look at the character table of the C_{2v} symmetry group, Tab. 6.2. We notice that the group has four (inequivalent) irreducible representations ($N_s = 4$) each of order $d_k = 1$. Consequently, the $\underline{\mathbf{T}}^{(k)}$ are all 1×1 matrices with the character $\chi_n^{(k)}$ as single element.

$\chi_n^{(k)}$	$\hat{e}$	$\hat{c}_2$	$\hat{\sigma}_x$	$\hat{\sigma}_y$
$\underline{\mathbf{T}}^{(1)}$	1	1	1	1
$\underline{\mathbf{T}}^{(2)}$	1	1	-1	-1
$\underline{\mathbf{T}}^{(3)}$	1	-1	-1	1
$\underline{\mathbf{T}}^{(4)}$	1	-1	1	-1

Table 6.2.: *Character table of the C_{2v} symmetry group [91].*

Following from the direct sum in Eq. (6.40), the representations in the field vector space with respect to the canonical basis are then given by matrices

$$\underline{\mathbf{T}}(\hat{e}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad \underline{\mathbf{T}}(\hat{\sigma}_x) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$
$$\underline{\mathbf{T}}(\hat{c}_2) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \qquad \underline{\mathbf{T}}(\hat{\sigma}_y) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \tag{6.44}$$

With these matrices, we have all ingredients to evaluate Eq. (6.41). Note, that the matrices are diagonal and, hence, the modes of a certain symmetry class do not mix with modes of another.

Let us now examine the case $n = 2$, $i = 3$ for the magnetic field $\mathbf{H}$, for instance. The position vector argument on the left hand side of Eq. (6.41) becomes

$$\mathbf{r}' = \underline{\mathbf{M}}_2^{-1}\,\mathbf{r} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -x \\ -y \\ z \end{pmatrix}. \tag{6.45}$$

And we get for the whole equation

$$\begin{aligned} \underline{\mathbf{M}}_2\,\mathbf{H}^{[3]}(\mathbf{r}') &= \sum_{j=1}^{4} \left(\underline{\mathbf{T}}_2\right)_{j3}\,\mathbf{\Psi}^{[j]}(\mathbf{r}) \\ &= 0\cdot\mathbf{H}^{[1]}(\mathbf{r}) + 0\cdot\mathbf{H}^{[2]}(\mathbf{r}) + (-1)\cdot\mathbf{H}^{[3]}(\mathbf{r}) + 0\cdot\mathbf{H}^{[4]}(\mathbf{r}) \\ \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} H_x^{[3]} \\ H_y^{[3]} \\ H_z^{[3]} \end{pmatrix}(\mathbf{r}') &= (-1)\cdot \begin{pmatrix} H_x^{[3]} \\ H_y^{[3]} \\ H_z^{[3]} \end{pmatrix}(\mathbf{r}). \end{aligned} \tag{6.46}$$

(x,y)	(-x,-y)	(x,-y)	(-x,y)
$E_x^{[1]}$ $H_x^{[1]}$	− −	+ −	− +
$E_y^{[1]}$ $H_y^{[1]}$	− −	− +	+ −
$E_z^{[1]}$ $H_z^{[1]}$	+ +	+ −	+ −
(m_1,m_2)	(-m_1,-m_2)	(m_1,-m_2)	(-m_1,m_2)

(x,y)	$(-x,-y)$	$(x,-y)$	$(-x,y)$
$E_x^{[2]}$ $H_x^{[2]}$	− −	− +	+ −
$E_y^{[2]}$ $H_y^{[2]}$	− −	+ −	− +
$E_z^{[2]}$ $H_z^{[2]}$	+ +	− +	− +
(m_1,m_2)	(-m_1,-m_2)	(m_1,-m_2)	(-m_1,m_2)

(x,y)	(-x,-y)	(x,-y)	(-x,y)
$E_x^{[3]}$ $H_x^{[3]}$	+ +	− +	− +
$E_y^{[3]}$ $H_y^{[3]}$	+ +	+ −	+ −
$E_z^{[3]}$ $H_z^{[3]}$	− −	− +	+ −
(m_1,m_2)	(-m_1,-m_2)	(m_1,-m_2)	(-m_1,m_2)

(x,y)	$(-x,-y)$	$(x,-y)$	$(-x,y)$
$E_x^{[4]}$ $H_x^{[4]}$	+ +	+ −	+ −
$E_y^{[4]}$ $H_y^{[4]}$	+ +	− +	− +
$E_z^{[4]}$ $H_z^{[4]}$	− −	+ −	− +
(m_1,m_2)	(-m_1,-m_2)	(m_1,-m_2)	(-m_1,m_2)

Table 6.3.: *Sign tables for the four C_{2v} symmetry modes. Explanation see text.*

From that, we deduce the following three *symmetry relations* between the field components of the magnetic field:

$$H_x^{[3]}(-x,-y) = +H_x^{[3]}(x,y)\,, \tag{6.47a}$$
$$H_y^{[3]}(-x,-y) = +H_y^{[3]}(x,y)\,, \tag{6.47b}$$
$$H_z^{[3]}(-x,-y) = -H_z^{[3]}(x,y)\,. \tag{6.47c}$$

Repeating this procedure for $n = 3$ and $n = 4$ ($n = 1$ is the identity relation) we get three more relations each, resulting in

$$H_x^{[3]}(x,y) = +H_x^{[3]}(-x,-y) = +H_x^{[3]}(x,-y) = +H_x^{[3]}(-x,y)\,, \tag{6.48a}$$
$$H_y^{[3]}(x,y) = +H_y^{[3]}(-x,-y) = -H_y^{[3]}(x,-y) = -H_y^{[3]}(-x,y)\,, \tag{6.48b}$$
$$H_z^{[3]}(x,y) = -H_z^{[3]}(-x,-y) = +H_z^{[3]}(x,-y) = -H_z^{[3]}(-x,y)\,. \tag{6.48c}$$

Similarly, we must repeat this procedure for the electric field and, in turn, the whole procedure for all other symmetry modes $i = 1, 2, 4$, as well.

The relations we obtain only differ in sign. We schematically listed these relations, or more precisely their signs, in Tab. 6.3. The symmetry relations of Eq. (6.48) are condensed in the lower left table with red symbols for the **H**-field components, for example. The signs of the electric field components are denoted with black symbols. To illustrate these rather unintuitive relations for each symmetry mode, we schematically depicted the field vectors, projected onto the x-y-plane, as arrows in Fig. 6.3. Filled (open) arrows mean a positive (negative) z-component. Blue arrows correspond to electric and red arrows to magnetic fields.

So far, we obtained the symmetry relations for the real-space fields, but for the reduction of the discretized eigenproblem we need the symmetry relations for the corresponding Fourier coefficients. In principle the symmetry relations also hold in **k**-space, at least as long as the reciprocal lattice reflects the same symmetry. The reciprocal lattice itself always possesses this symmetry because

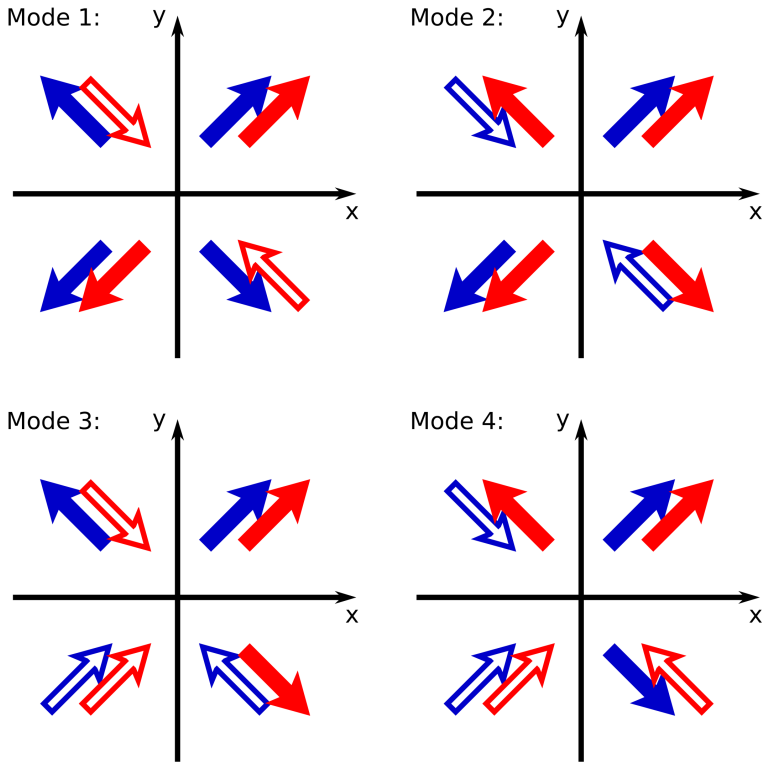


Figure 6.3.: *Schematic illustration of the four C_{2v} symmetry modes in real-space. Depicted by the arrows are the projections of the fields into the x-y-plane. Blue (red) arrows correspond to electric (magnetic) fields. Filled (open) arrows mean a positive (negative) z-component.*

it inherits it by construction from the direct lattice. However, the reciprocal lattice is in general displaced from the origin by the transverse part of the incoming light's wave vector. There are a few mountings of the incident wave, when the symmetry is restored — called Lithrow mountings. This occurs when the in-plane wave vector components α_0 and β_0 of the incoming wave (cf. Eq. (3.44)) are either integer or half-integer multiples of the reciprocal lattice vectors. Still, in these cases the incident field must first be symmetrized by an expansion into its canonical basis as discussed above. For a detailed discussion we refer the reader to Ref. [91]. Here, we restrict our considerations to the case of normal incidence, where the symmetry of the reciprocal lattice is evident — waveguide eigenmodes have no in-plane wave vector components as their propagation direction is along the fiber axis.

We examine the real-space symmetry relation $H_x^{[3]}(x,y) = +H_x^{[3]}(-x,-y)$. Expanding both sides in a truncated Floquet-Fourier series (cf. Eq. (6.4)) we obtain

$$\begin{aligned}\sum_{m=1}^{M} \tilde{\mathsf{H}}_{x,m}^{[3]}\, e^{\mathrm{i}(\alpha_{m_1}x+\beta_{m_2}y)} &= +\sum_{n=1}^{M} \tilde{\mathsf{H}}_{x,n}^{[3]}\, e^{\mathrm{i}(-\alpha_{n_1}x-\beta_{n_2}y)} \\ &= +\sum_{n=1}^{M} \tilde{\mathsf{H}}_{x,n}^{[3]}\, e^{\mathrm{i}(\alpha_{-n_1}x+\beta_{-n_2}y)}\,. \end{aligned} \tag{6.49}$$

A comparison of Fourier coefficients leads to their symmetry relation

$$\tilde{\mathsf{H}}^{[3]}_{x,m_1m_2} = +\tilde{\mathsf{H}}^{[3]}_{x,-m_1-m_2}\,, \tag{6.50}$$

which has the same sign as the symmetry relation of the real-space fields. This also holds true for all other symmetry relations, and we can use Tab. 6.3 for the Fourier coefficients considering the column labels in the bottom lines, as well. Furthermore, we can deduce that some Fourier coefficients on the axes must be zero, e.g.,

$$E^{[1]}_{x,m_1m_2} = -E^{[1]}_{x,-m_1m_2} \quad \xrightarrow{m_1=0} \quad E^{[1]}_{x,0m_2} = -E^{[1]}_{x,0m_2} = 0\,. \tag{6.51}$$

Thus, if we consider all symmetry relations, the minimal set of Fourier coefficients, which we have to take into account for the eigenproblem to retain all information, are those in the first $\mathbf{k}$-space quadrant enclosed in the boxed regions in Fig. 6.4. The minimal sets are different for the field components

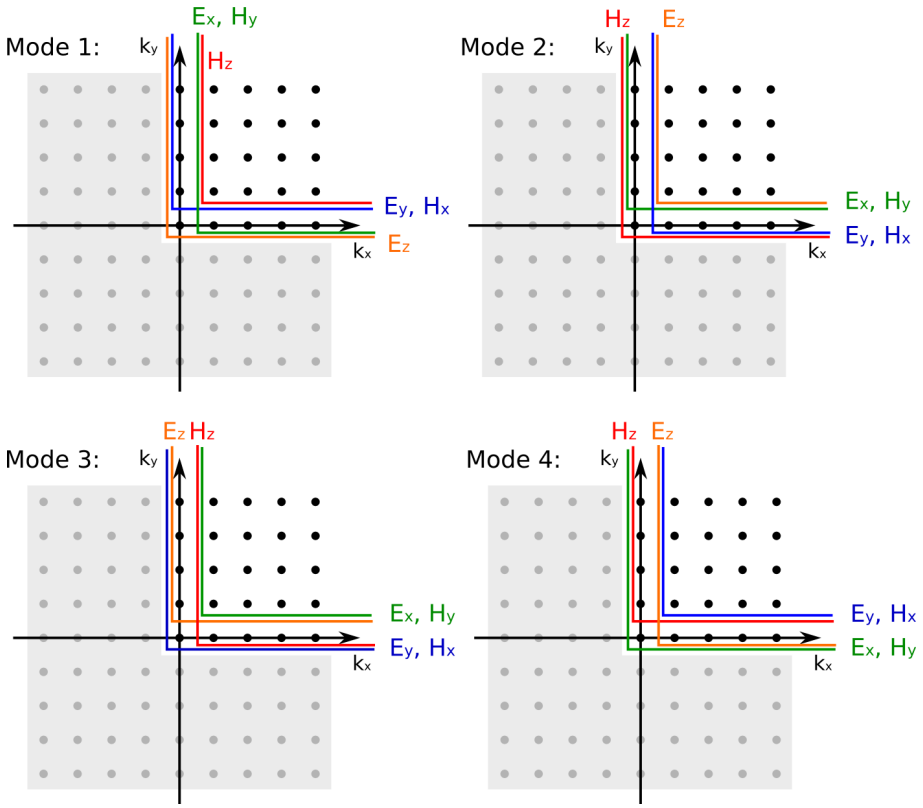


Figure 6.4.: *Reduction of the considered k-space due to C_{2v}-symmetry. The Fourier coefficients corresponding to the lattice points in the regions up and right of of the colored borders are the minimal set that still contains the information of the full problem. The retained coefficients depend on symmetry mode and field component. The latter is highlighted by the used colors.*

(colored boxes) and depend on the symmetry mode. All other Fourier coefficients are either zero (those on the axes) or can be regained by symmetry relations.

7 Chapter 7.
Coordinate Transformations

The Fourier modal method, as introduced in Chap. 6, works on equidistant Cartesian grids (cf. Fig. 7.1(a)). These grids are required — because the Fourier transformation is achieved with the efficient FFT algorithm — but they are also a huge restriction to applicability and efficiency of the method. Some modifications to the basic algorithm have been introduced in the literature [75] in order to be able to simulate structures with non-orthogonal in-plane lattices, e.g., hexagonal lattices. Recalling that we choose the first lattice vector $\mathbf{a}_1$ to always coincide with the first basis vector $\mathbf{e}_1$ of the Cartesian coordinate system, the second lattice vector $\mathbf{a}_2$ must then include an angle with the second basis vector $\mathbf{e}_2$ — the lattice angle ξ. The fact that the second lattice vector is not aligned to the basis vectors leads to additional terms in the eigenvalue equation, which involve sine or cosine

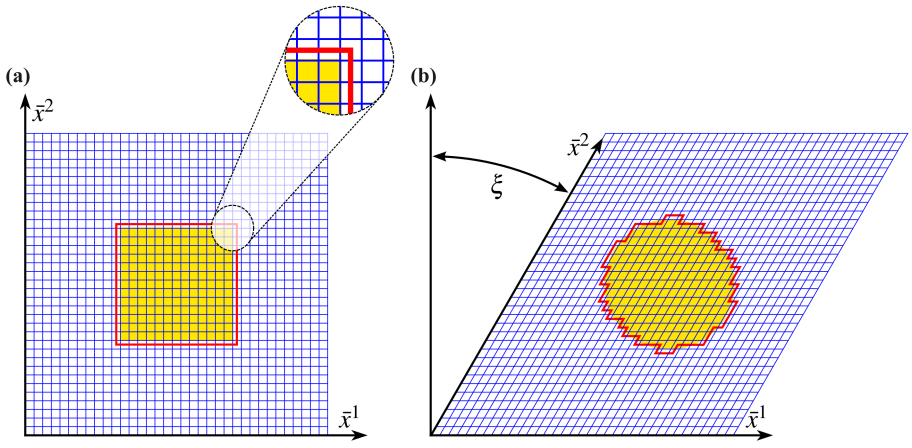


Figure 7.1.: *Illustration of regular equidistant meshes (blue): (a) Cartesian mesh and (b) tilted mesh. The left panel shows a grid-aligned square structure and the right panel a circular structure (yellow). The circular structure is staircased during the numerical discretization. The surfaces of the discretized structures are highlighted in red. Depending on the number of sampling points, the numerical surface does not necessarily coincide with the physical surface even for a grid-aligned structure (inset).*

terms of the lattice angle. As a consequence, the eigenvalue equation must be adapted to the problem. The discretization of the unit cell is then performed on a tilted equidistant mesh as can be seen in Fig. 7.1(b).

Such modifications improve the applicability of the FMM to systems with non-orthogonal lattices. Still, there remain several other problems, the most prominent of which is a suboptimal representation of structure surfaces due to in-plane staircasing caused by the finite discretization. This effect is also illustrated in Fig. 7.1. The physical structure is numerically represented by permittivity and permeability values sampled on the equidistant mesh, where each sampling point is associated with a finite (square or parallelogrammic) elemental area. Assigning a permittivity (permeability) value to a sampling point implicitly assigns it to the whole elemental area. The resulting numerical structure differs from the physical structure quite a bit — at least as long as the structure is not grid-aligned.[1] Round surfaces are represented angularly with steps. The deviation of the sampled structure from the analytic structure can be reduced by increasing the number of sampling points. However, the staircasing of non-grid aligned surfaces is always present and thus a finite deviation remains. Furthermore, it is well known that electromagnetic fields at sharp metallic corners tend to diverge [19, 52, 96] — leading to the so-called lightning rod effect.

Another undesired effect of equidistant meshes is that the resolution is everywhere the same — even in regions where neither structure nor electromagnetic fields vary much. In such regions the high resolution is overkill whereas at material surfaces, where the material parameters have jump discontinuities, the resolution is usually too low to represent the functions with a finite number of Fourier coefficients satisfactorily (cf. Gibbs' phenomenon Sec. 3.4.3). Numerical methods working with unstructured meshes are often capable of a local resolution enhancement in areas where interesting physical effects are expected to happen. Since the manageable degrees of freedom in the FMM are limited in practical applications because the eigenproblem scales unfavorably, this drawback is a crucial point especially for three-dimensional problems.

The concept of *coordinate transformations* provides an opportunity to partially overcome these obstacles and profoundly increase the performance of the FMM. The key words are *adaptive coordinates* (AC) and *adaptive spatial resolution* (ASR). The former describes the adaption of the coordinate system to the investigated structure by setting up coordinate lines parallel to material surfaces. This helps to avoid in-plane staircasing and improves the representation's accuracy. The latter describes the additional local increase of the coordinate line density which helps to improve the convergence rate of the Fourier series [97, 98]. Among others, such regions include the vicinity of jump discontinuities of permittivity and permeability where the Fourier representation is generally difficult (cf. Sec. 3.4.3).

The incorporation of coordinate transformations into the Fourier modal method is possible in two different ways. In the first approach, the coordinate transformations enter Maxwell's equations directly and, thus, modify the eigenproblem equation. The eigenproblem then fundamentally depends on the used coordinate system. We call this the *equation-transform k-space strategy*. In the second approach, the eigenproblem stays unaltered. Instead, the structure, i.e., the material parameters $\underline{\varepsilon}(\mathbf{r})$ and $\underline{\mu}(\mathbf{r})$, is modified such that the problem can be conveniently calculated with the FMM algorithm. Afterwards, the modification is reversed for the calculated solutions. For this approach we

[1] The term grid-aligned describes that the physical surfaces are all parallel to the coordinate lines of the grid, e.g., an unrotated square is grid-aligned to a Cartesian mesh. Even grid-aligned structures are not necessarily represented, because the sampling points next to the surface may not be placed symmetric to it.

coin the name *structure-transform real-space strategy*. By virtue of the concept of transformation optics, both approaches are equivalent as long as some requirements for the structure transform are met: The covariant formulation of Maxwell's equations must be used, the structure modification is done according to a proper coordinate transformation, and the solutions are transformed back into the original coordinate system in the end. The covariant formulation and coordinate transformations have already been introduced in Sec. 2.5. The structure transform variant is much more flexible and easier to incorporate into the existing FMM algorithm because it leaves the eigenproblem invariant, which is the reason why we predominantly favor this approach. One of the drawbacks of coordinate transformations is that even isotropic structures are transformed into anisotropic structures. This is one of the reasons why we kept the tensorial structure of both material parameters throughout this work. Adaptive coordinates and adaptive spatial resolution are topic of Sec. 7.1.

A further advantage of this approach is that not only real coordinate mappings of finite unit cell regions can be performed with this machinery. Instead, it allows for coordinate mappings of infinitely extended complex regions onto a finite unit cell as well, which enables us to simulate open boundary conditions even in a periodic setup inevitable for the method's underlying Floquet-Fourier expansion. These open boundary conditions — commonly called (stretched-coordinate) *perfectly matched layers* (PML) — have first been reported (using the *equation-transform k-space strategy*) for the two-dimensional standard FMM by Lalanne *et al.* [99]. We discuss PMLs in Sec. 7.2 and generalize them to full three-dimensional problems.[2] In Sec. 7.4, we present both the old *equation-transform k-space strategy* and the new *structure-transform real-space strategy* in the context of PMLs in more details, and highlight their numerical differences.[3]

7.1. Adaptive Coordinates and Adaptive Spatial Resolution

The crucial part of the procedure is the construction of an adapted mesh which improves the representation of the structure. We would like to emphasize that not only an improved representation in direct space is desirable, but also an improvement in Fourier space. The perfect representation in direct space is achieved when the sampled structure exactly matches the original structure. The perfect improvement in Fourier space would be achieved if the coordinate transformation modified the material parameters to bandwidth limited functions.[4] Because of the usual jump discontinuities at material interfaces this seems impossible. Consequently, the best one can achieve is a fast decay of Fourier coefficients away from the origin.[5] The faster this decay, the better we expect the performance of the simulation.

As a rule of thumb, the AC mesh adaption to the material surfaces is mainly responsible for an accurate direct space representation, whereas the ASR predominantly contributes to a convenient Fourier space representation.

[2] This, to the best of our knowledge, is the first time 3D PMLs in FMM have ever been rigorously written down.

[3] Before we proceed, we would like to mention that this chapter has a large overlap with Ref. [100], which was a collaborative work with J. Küchenmeister. The author's contributions to this reference comprise: Initial idea for smoothed meshes, considerable contributions to their conceptional design, C++ coding, implementation of reference solutions, and major contributions to the theoretical data analysis.

[4] Bandwidth limited functions are characterized by vanishing Fourier coefficients beyond a certain order.

[5] In typical three-dimensional simulations the highest retained order is approximately between $\sqrt{M/\pi} \approx 15 \ldots 25$ in circular truncation (cf. Sec. 3.4.2).

An adapted mesh simultaneously defines a curvilinear coordinate system, where the basis vectors are tangential to the lines of the adapted mesh. This implies that there exists a coordinate transformation between the new adapted mesh, which is given with respect to the Cartesian coordinate system, and its corresponding Cartesian mesh[6] in the curvilinear adapted coordinate system. This coordinate transformation describes the mapping between both coordinate systems.

Recalling Sec. 2.5.2, we define the Cartesian coordinate system $\mathcal{O}\bar{x}^1\bar{x}^2\bar{x}^3$ with barred quantities, and the curviliniear coordinate system $\mathcal{O}x^1x^2x^3$ with unbarred quantities. The coordinate transformation itself is described by functions of the form

$$\bar{x}^1 = \bar{x}^1(x^1, x^2, x^3)\,, \tag{7.1a}$$
$$\bar{x}^2 = \bar{x}^2(x^1, x^2, x^3)\,, \tag{7.1b}$$
$$\bar{x}^3 = \bar{x}^3(x^1, x^2, x^3)\,, \tag{7.1c}$$

given in the Cartesian coordinate system. Even though the procedure is in general equally well applicable to three-dimensional coordinate transformations, we restrict the examples in this work to coordinate transformations in the transverse plane within the layers.[7] This means Eqs. (7.1) reduce to

$$\bar{x}^1 = \bar{x}^1(x^1, x^2)\,, \tag{7.2a}$$
$$\bar{x}^2 = \bar{x}^2(x^1, x^2)\,, \tag{7.2b}$$
$$\bar{x}^3 = x^3\,, \tag{7.2c}$$

where the x^3 coordinates along the primary axis remain untransformed.

The structure in the curvilinear coordinate system in the l-th layer is characterized by the *effective permittivity* $\underline{\boldsymbol{\varepsilon}}^{(l)}(x^1, x^2)$ and *effective permeability* $\underline{\boldsymbol{\mu}}^{(l)}(x^1, x^2)$ (cf.Eqs. (2.86)) with tensor elements

$$\varepsilon^{\rho\sigma} = \sqrt{g}\,\frac{\partial x^\rho}{\partial \bar{x}^\tau}\frac{\partial x^\sigma}{\partial \bar{x}^\kappa}\,\bar{\varepsilon}^{\tau\kappa}\,, \tag{7.3a}$$
$$\mu^{\rho\sigma} = \sqrt{g}\,\frac{\partial x^\rho}{\partial \bar{x}^\tau}\frac{\partial x^\sigma}{\partial \bar{x}^\kappa}\,\bar{\mu}^{\tau\kappa}\,, \tag{7.3b}$$

where $\bar{\varepsilon}^{\tau\kappa}$ and $\bar{\mu}^{\tau\kappa}$ denote the respective material function tensor components in the Cartesian space. Again, we omit the layer labels where it is clear from the context. We recall that the metric tensor determinante is denoted by

$$g = \det \underline{\boldsymbol{G}} \tag{7.4a}$$

with metric tensor components (cf. Eq. (2.88))

$$\left(\underline{\mathbf{G}}\right)_{\rho\sigma} = g_{\rho\sigma} = \frac{\partial \bar{x}^\tau}{\partial x^\rho}\frac{\partial \bar{x}^\tau}{\partial x^\sigma}\,. \tag{7.4b}$$

[6] Which is essential to the FMM.

[7] Numerical experiments with 3D coordinate transformations have already been performed in collaboration with J. Küchenmeister. However, this work is still in progress and a few issues remain to be solved.

Since the original problem is given in Cartesian space with unit metric $\underline{\bar{\mathbf{G}}} = \underline{\mathbf{1}}$, we get $\sqrt{\bar{g}} = 1$ (cf. Eq. (7.4a)). According to Eq. (2.79), the prefactor then equals the determinant of the transformation matrix

$$\sqrt{g} = \det \underline{\mathbf{\Lambda}} \,. \tag{7.5}$$

The transformation matrix is usually called the Jacobian matrix.

In most cases it is much easier to write down the coordinate transformation functions as in Eqs. (7.2) rather than its inverse functions $x^\rho(\bar{x}^\sigma)$. However, in Eqs. (7.3) we need the curvilinear coordinates' derivatives with respect to the Cartesian coordinates. Therefore, it is quite convenient to express these derivatives in terms of derivatives of Cartesian coordinates with respect to the curvilinear coordinates. Using the inverse function theorem for the Jacobian matrices [101] and Cramer's rule [28, 102], we explicitly obtain for in-plane transformations (cf. Eq. (2.73)):

$$\begin{pmatrix} \frac{\partial x^1}{\partial \bar{x}^1} & \frac{\partial x^1}{\partial \bar{x}^2} & 0 \\ \frac{\partial x^2}{\partial \bar{x}^1} & \frac{\partial x^2}{\partial \bar{x}^2} & 0 \\ 0 & 0 & 1 \end{pmatrix} = \underline{\bar{\mathbf{\Lambda}}} \quad \stackrel{!}{=} \quad \underline{\mathbf{\Lambda}}^{-1} = \begin{pmatrix} \frac{\partial \bar{x}^1}{\partial x^1} & \frac{\partial \bar{x}^1}{\partial x^2} & 0 \\ \frac{\partial \bar{x}^2}{\partial x^1} & \frac{\partial \bar{x}^2}{\partial x^2} & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1} = \frac{1}{\det \underline{\mathbf{\Lambda}}} \begin{pmatrix} \frac{\partial \bar{x}^2}{\partial x^2} & -\frac{\partial \bar{x}^1}{\partial x^2} & 0 \\ -\frac{\partial \bar{x}^2}{\partial x^1} & \frac{\partial \bar{x}^1}{\partial x^1} & 0 \\ 0 & 0 & \det \underline{\mathbf{\Lambda}} \end{pmatrix} . \tag{7.6}$$

On the one hand, we can easily read of the Jacobian determinant from the center matrix, which is given by

$$\det \underline{\mathbf{\Lambda}} = \left(\frac{\partial \bar{x}^1}{\partial x^1} \frac{\partial \bar{x}^2}{\partial x^2} - \frac{\partial \bar{x}^2}{\partial x^1} \frac{\partial \bar{x}^1}{\partial x^2} \right) \stackrel{Eq.\ (7.5)}{=} \sqrt{g} \,. \tag{7.7}$$

On the other hand, Eq. (7.6) provides the desired replacements. Hence, the derivatives in Eqs. (7.3) can be conveniently replaced with expressions which are directly accessible.

The necessary derivatives are numerically carried out on the material's discretized real-space representations. Considering an equidistant sampling of the transformed problem as described by Eqs. (3.21), we numerically carry out derivatives as

$$\left. \frac{\partial \bar{x}^\rho}{\partial x^1} \right|_{k,l} = \frac{-\bar{x}^\rho(k+2,l) + 8\bar{x}^\rho(k+1,l) - 8\bar{x}^\rho(k-1,l) + \bar{x}^\rho(k-2,l)}{12\Delta_{x^1}} + \mathcal{O}\left((\Delta_{x^1})^4 \right) , \tag{7.8a}$$

$$\left. \frac{\partial \bar{x}^\rho}{\partial x^2} \right|_{k,l} = \frac{-\bar{x}^\rho(k,l+2) + 8\bar{x}^\rho(k,l+1) - 8\bar{x}^\rho(k,l-1) + \bar{x}^\rho(k,l-2)}{12\Delta_{x^2}} + \mathcal{O}\left((\Delta_{x^2})^4 \right) , \tag{7.8b}$$

where indices k and l denote the considered sampling point with coordinates (x^1_k, x^2_l) given by the mesh in the adapted space and its corresponding Cartesian coordinates $\bar{x}^\rho(k,l) = \bar{x}^\rho(x^1_k, x^2_l)$.

At this point we would like to stress that the whole transformation process could be done analytically if we used analytic coordinate transformations only. In our code the discretization is never performed in Cartesian space but solely in transformed space according to the equidistant rectangular mesh obligatory for the FFT. The discretization in Cartesian space then directly follows by means of the transformed coordinates. The use of numerical derivatives in the transformation as described by Eqs. (7.8) is for convenience and generality (non-analytic coordinate transformation, see Sec. 7.1.2) whilst our studies show little influence on the overall accuracy.

As stated before, Maxwell's equations as used in the derivation of the eigenproblem in Eqs. (5.4) stay the same in the new adapted coordinate system (they are covariant). This means we simply

replace the Cartesian material functions in the eigenproblem operator, Eq. (5.8), with the modified curvilinear material functions provided by Eqs. (7.3) together with Eq. (7.6) and Eq. (7.7).[8] This is everything that needs to be done to reformulate the problem in the new coordinate system. The whole FMM machinery can be used without further modifications. Only in the end we will transform back the solutions (fields) where necessary for convenience of an intuitive interpretation.

The remaining task is the construction of proper adapted meshes.[9] This means we have to find suitable expressions for Eqs. (7.2). We can either write them down analytically (cf. Sec. 7.1.1 Analytic Adapted Mesh Construction), or use a computational algorithm to calculate them numerically (cf. Sec. 7.1.2 Automated Adapted Mesh Generation).

Before we present these approaches in the subsequent sections in more detail, it might be instructive to visualize the general idea in more detail with the help of an example [100]. The permittivity distribution of a circular structure within the unit cell of a quadratic lattice may look like depicted in Fig. 7.2(a). The plot depicts the permittivity distribution $\bar{\varepsilon}(\bar{x}^1, \bar{x}^2)$ of a circular structure of radius $r = 0.3a$ made of dielectric material with $\bar{\varepsilon}_{struc} = 2$ centered in the square unit cell of size a with a background material $\bar{\varepsilon}_{bg} = 1$. The used materials are isotropic. The permittivity is discretized on an adapted mesh with 1024×1024 sampling points throughout the unit cell, which is the standard resolution we use in FMM calculations.[10]

The corresponding adapted mesh is presented in Fig. 7.2(b). It is an analytically constructed non-differentiable mesh discussed in Refs. [100, 103]. The plot shows 80 out of the 1024 coordinate lines per dimension. The sector in the red box is zoomed in on the right hand side. The mesh features two important properties: First, there are many points where the coordinate lines are not differentiable. Some of them are marked with green, dashed circles. Second, the red circle marks one of the four points in the mesh where the coordinate lines of the mesh run parallel in $\bar{x}^1$ and $\bar{x}^2$ direction.

The effective permittivity tensor $\underline{\boldsymbol{\varepsilon}}(x^1, x^2)$ resulting from the depicted coordinate transformation is plotted in Fig. 7.2(c). The 3×3 array of plots in this panel show the permittivity distribution throughout the unit cell for every tensor component $\varepsilon^{\rho\sigma}$ (from left to right, from top to bottom: ε^{11}, ε^{12}, ε^{13}, ε^{21}, ...). Those are the quantities that are actually Fourier transformed when using FMM with adaptive coordinates. We would like to emphasize a couple of typical features: First of all, the circular shape has transformed into a quadratic shape. Most importantly, discretizing this new grid-aligned structure can be done much more precisely than before. The staircased (discretized) surface can easily be chosen to coincide with the natural surface of the structure. As a consequence, Li's one-dimensional factorization rules can be effectively applied to the new grid-aligned material interface. Secondly, we immediately notice the price we have to pay for a grid-aligned structure besides the already mentioned anisotropy — depending on the tensor component under investigation, the non-differentiable points (c.f. green circles in Fig. 7.2(b)) in the mesh lead to additional discontinuities in the effective permittivity. Last but worst, the permittivity values in the corners of the square diverge, which is not visible in the plot because the color scale is saturated at five. Furthermore, in practical applications the finite discretization prevents the numerical values from approaching infinity. In the present example the maximum value is about 5000. The reason for this undesired and unfavorable

[8] Still, we must apply Li's operator to obtain $\check{\underline{\boldsymbol{\varepsilon}}} = \hat{1}_3^-(\underline{\boldsymbol{\varepsilon}})$.

[9] Note, that an adapted mesh is the coordinate transformed equidistant (Cartesian) mesh in the curvilinear space. See Sec. 3.4.2.

[10] The staircasing steps that may be visible on the surface of the circular structure are purely due to the rasterization of the image.

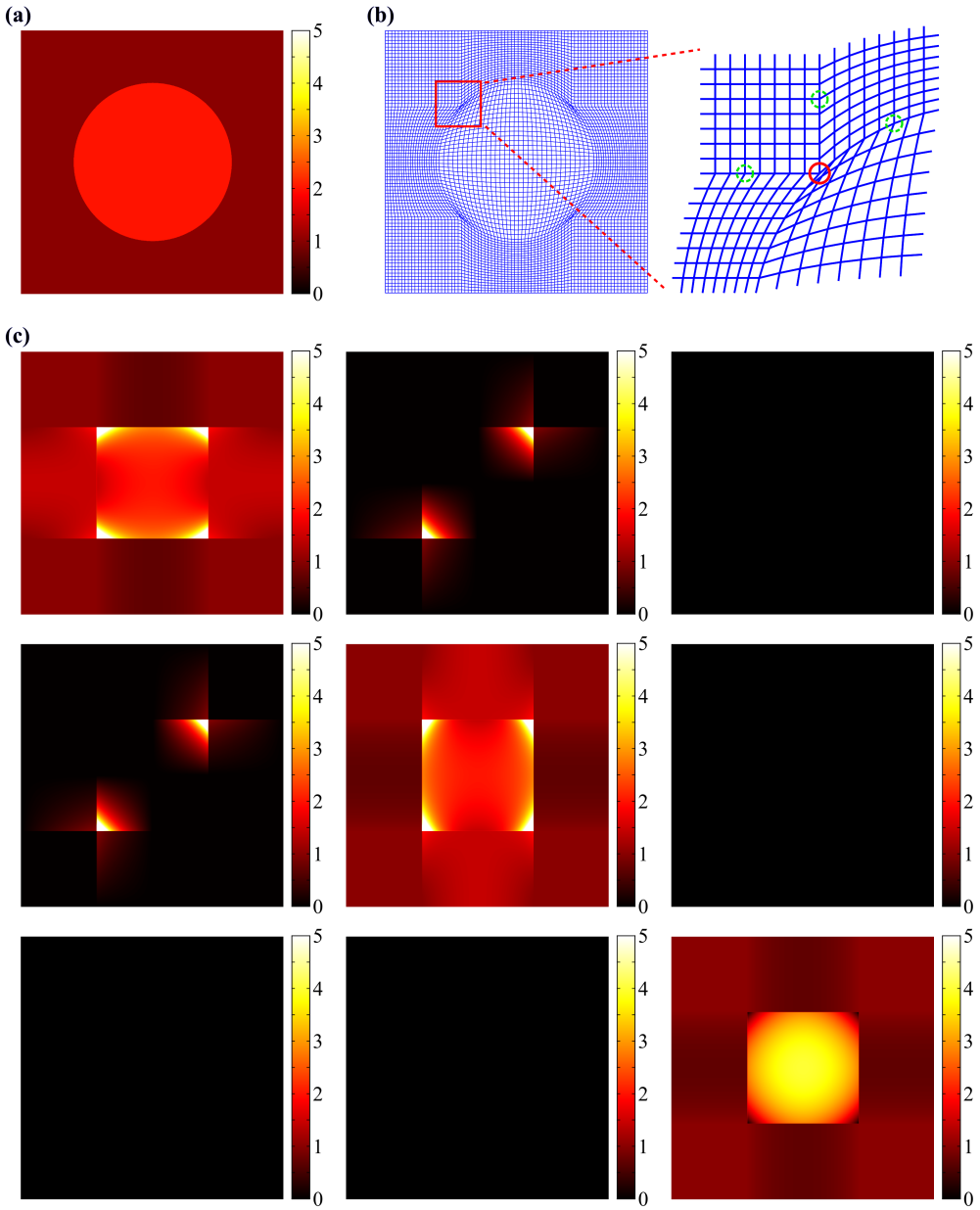


Figure 7.2.: *Illustration of the effect of adaptive coordinates to the material functions. Panel (a) depicts the isotropic permittivity distribution $\bar{\varepsilon}(\bar{x}^1, \bar{x}^2)$ of a circular dielectric structure with radius $r = 0.3$ centered in a square unit cell in Cartesian space. Panel (b) shows the used adapted mesh as introduced in Refs. [100, 103], and a close-up of the region marked with the red box. Some typical non-differentiable points have been highlighted with green, dashed circles. The red circle marks one point where the coordinate lines of the mesh run parallel in $\bar{x}^1$ and $\bar{x}^2$ direction. The 3×3 array of plots in panel (c) illustrates the distribution of the full anisotropic effective permittivity tensor $\underline{\boldsymbol{\varepsilon}}(x^1, x^2)$ in the corresponding transformed curvilinear space. All color scales are equal and cut at a value of five in order to improve the visibility of details. The in-plane components exceed this limit by far and reach values up to 5000. All distribution plots show 1024×1024 data points. See text for more details.*

singular behavior becomes clear when we have a close look at Eq. (7.6) and Eq. (7.7). At points where coordinate lines of directions $\bar{x}^1$ and $\bar{x}^2$ are parallel (cf. red circle in Fig. 7.2(b)), i.e.,

$$\frac{\partial \bar{x}^1}{\partial x^1} = \frac{\partial \bar{x}^2}{\partial x^1} \qquad \text{and} \qquad \frac{\partial \bar{x}^2}{\partial x^2} = \frac{\partial \bar{x}^1}{\partial x^2} \,, \tag{7.9}$$

the Jacobian determinant vanishes. This leads to diverging derivatives $\partial x^\rho / \partial \bar{x}^\sigma$ and, thereby, to a diverging effective permittivity. We would like to mention that the permeability of the problem looks qualitatively the same as the permittivity. The only difference is that the effective permeability within and outside the square is reduced by the factor of the respective permittivity within and outside the circle ($\bar{\mu} = 1$ everywhere instead of $\bar{\varepsilon}_{\mathrm{struc}}$ and $\bar{\varepsilon}_{\mathrm{bg}}$).

7.1.1. Analytic Adapted Mesh Construction

We start the discussion of analytic adapted meshes by pointing out that there are two distinct interpretations that may be associated with the coordinate transformations discussion in general [100]. The first interpretation is that we map a given permittivity distribution onto a new distribution according to a given mesh. This means that we go from bent coordinate lines in Cartesian space to straight coordinate lines in the transformed space. This interpretation is particularly useful when we discuss the shape of the transformed material matrix in the FMM as in the example of Fig. 7.2. The second interpretation is very handy for the mesh construction, which we are up to in the following. Here, we proceed the other way round and figure out how to map straight coordinate lines onto bent lines that match the surface given by the material distribution. Both these interpretations are valid — which one we use depends on whether the construction or the application of the meshes is emphasized.

In this section we would like to give a short overview of the principles of analytic adapted mesh construction. Analytic meshes in our understanding are meshes obtained from an analytic coordinate transformation — a coordinate transformation that can be written down in a closed form throughout the whole unit cell. So what we really do when we talk about analytic mesh construction is writing down an analytic coordinate transformation

Of course, for every coordinate transformation there exists an infinite number of meshes distinct by the associated space discretization (mesh) in the transformed space. We only consider regular equidistant Cartesian meshes in transformed space. The number of discretization points per dimension we leave as a parameter. Hence, the concrete analytic adapted mesh is determined by the coordinate transformation functions and the numbers of discretization points for both transversal dimensions.

The primary goal is to achieve grid-aligned structure surfaces in the transformed space. Consequently, either a

$$\bar{x}^1\text{-coordinate line} \qquad \text{with} \qquad \left(\bar{x}^1, \bar{x}^2\right) : x^1 = \text{const.} \tag{7.10a}$$

or a

$$\bar{x}^2\text{-coordinate line} \qquad \text{with} \qquad \left(\bar{x}^1, \bar{x}^2\right) : x^2 = \text{const.} \tag{7.10b}$$

should run parallel to the material surface — or rather exactly on the surface. Several of these coordinate lines together cover the whole surface. These coordinate lines on structure surfaces we call

the specific lines. Their crossings define the characteristic points. Specific lines and characteristic points are the basic elements on which we build the construction of the corresponding coordinate transformation.

The general mesh construction procedure is as follows:

1. Select characteristic points on the structure surface. Usually two points per dimension per surface are required. The points define the specific coordinate lines in the transformed space. The specific lines and the outer boundaries divide the whole unit cell into several domains.
2. Parametrize the surface between the characteristic points to obtain analytic expressions for the mapping of the specific lines onto the surface.
3. Each domain is mapped by linear interpolation between their limiting specific lines and/or unit cell boundaries.

The linear interpolation between the specific lines is performed using the handy linear transition function

$$\bar{x} = LT(a, \bar{a}, b, \bar{b}, x) = \underbrace{\frac{(\bar{b} - \bar{a})}{(b - a)}}_{M} \cdot x + \underbrace{\bar{a} - a\frac{(\bar{b} - \bar{a})}{(b - a)}}_{\bar{x}_0}, \tag{7.11}$$

which defines a straight line through the points $(a, \bar{a})$ and $(b, \bar{b})$ (cf. Fig. 7.3(a)) [100]. As illustrated in Fig. 7.3(b), we can use the LT function to map whole coordinate lines. The regions between specific lines in transformed and Cartesian space are mapped by substituting those specific lines into the LT function replacing lines a, b and $\bar{a}$, $\bar{b}$, respectively. The coordinate lines of the adapted mesh can be obtained by evaluation of the resulting LT function at a set of equidistant values x.

Non-Differentiable Meshes

For the purpose of a simple illustration, we stick to the example of the circular structure's mesh used in Fig. 7.2. The application of the basic construction principles to more advanced structures can be found in Ref. [100]. Figure 7.4 shows the specific lines (blue: specific $\bar{x}^1$-lines, red: specific $\bar{x}^2$-

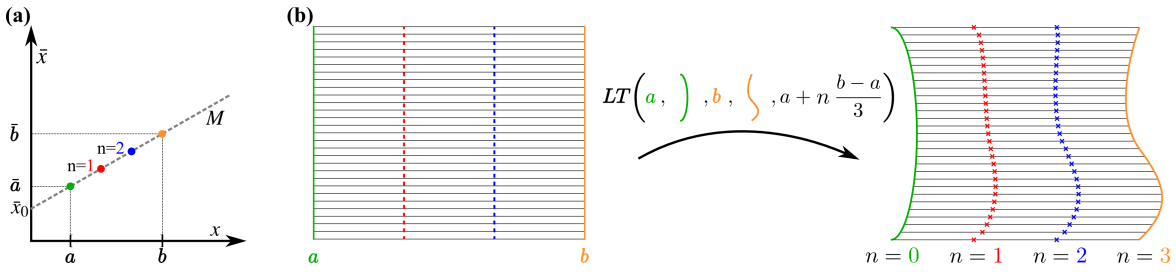


Figure 7.3.: *(a) Sketch of the LT function. (b) Application of the LT function to map intermediate coordinate lines. The straight lines a and b (specific lines in transformed space) are mapped onto the respective green curve and orange curve (specific lines in Cartesian space) on the right hand side. The linear transformation maps the equidistant red and blue dashed intermediate lines to the coordinates marked by red crosses and blue crosses, respectively.*

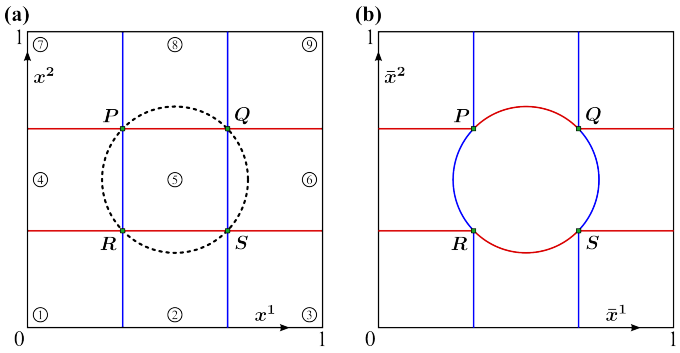


Figure 7.4.: *Specific lines and characteristic points (a) in transformed space and (b) in Cartesian space for a non-differentiable mesh of a circular structure. Panel (b) reproduced from [100].*

lines) and characteristic points (green) of the circular structure in both spaces.[11] The characteristic points define where the structure's surface parametrization switches from a $\bar{x}^1$- to a $\bar{x}^2$-line or vice versa. They can be placed nearly anywhere on the surface — with the constraint that we must obey periodic boundary conditions. Here, they are chosen most symmetrically on the diagonals of the unit cell. For more complex structures, the characteristic points are usually chosen such that the surface parametrization is easiest.

The parametrization of the surface by the specific $\bar{x}^1$- and $\bar{x}^2$-lines provides us with analytical expressions. In the circular structure case they are simple circle arcs, where the circle's center coincides with the center of the unit cell at the point (0.5,0.5). The corresponding coordinates can be conveniently represented by circle arc (CA) functions [100]

$$\begin{aligned} CA_{L/R}(x^2) &= 0.5 \mp \sqrt{r^2 - \Big(x^2 - 0.5\Big)^2}\,, \\ CA_{T/B}(x^1) &= 0.5 \pm \sqrt{r^2 - \Big(x^1 - 0.5\Big)^2}\,. \end{aligned} \tag{7.12}$$

The subscripts L,R,T, and B refer to left, right, top, and bottom pieces, respectively. The specific lines divide the unit cell into nine regions, which can be mapped independently. The corner regions ①, ③, ⑦, and ⑨ are just the identity transformations. Regions ②, ④, ⑥, and ⑧ are almost equivalent, which is why we pick region ② for the demonstration only. In this region the first coordinate is unchanged

$$\bar{x}^1(x^1, x^2) = x^1\,, \qquad \text{for } (x^1, x^2) \in ②\,, \tag{7.13a}$$

and the second coordinate is given by

$$\bar{x}^2(x^1, x^2) = LT\Big(\underbrace{0 \quad , \quad 0}_{\substack{\text{map lower edge} \\ \text{onto intself}}}, \underbrace{x^2_{RS}, CA_B(x^1)}_{\substack{\text{map RS line section} \\ \text{onto bottom circle arc}}}, x^2\Big)\,, \qquad \text{for } (x^1, x^2) \in ②\,. \tag{7.13b}$$

Similar expressions can be noted down for the equivalent regions.

[11] For ease of complexity, we assume that the characteristic points remain at the same positions in both coordinate systems.

Last but not least, the center region ⑤ follows analogously. We map in x^1-direction the PR-line to the left circle arc and the QS-line to the right circle arc

$$\bar{x}^1(x^1, x^2) = LT\Big(x^1_{PR}, CA_L(x^2), x^1_{QS}, CA_R(x^2), x^1\Big)\,, \qquad \text{for } (x^1, x^2) \in ⑤\,, \tag{7.14a}$$

and in x^2-direction the RS-line to the bottom circle arc and the PQ-line to the top circle arc

$$\bar{x}^2(x^1, x^2) = LT\Big(x^2_{RS}, CA_B(x^1), x^2_{PQ}, CA_T(x^1), x^2\Big)\,, \qquad \text{for } (x^1, x^2) \in ⑤\,. \tag{7.14b}$$

The described construction principle presents a general approach to analytic mesh construction and works for a very large variety of structures. It is noteworthy that the original derivation of non-differentiable meshes for circular structures was published by Weiss et al. [103]. However, the presented scheme (cf. Ref. [100]) is much clearer, easily applicable to other structures, and extendible. The following construction schemes mainly differ in the definition of the characteristic lines.

Smoothed Meshes

The meshes constructed with the non-differentiable scheme contain points where they are non-differentiable and where coordinate lines in $\bar{x}^1$- and $\bar{x}^2$-direction run in parallel, as already mentioned in the discussion of Fig. 7.2(b). The diverging effective permittivity in the corners of the transformed structure was attributed to the latter. *Smoothed meshes* are constructed in such a way as to avoid these parallel coordinate lines. This is achieved by smoothing the specific coordinate lines in Cartesian space at the junctures between circle arc and straight lines. As a consequence, the specific $\bar{x}^1$ and $\bar{x}^2$ coordinate lines always cross under a non-zero (or non-pi) angle. While the smoothness of the transition function could be optimized with splines or Bézier curves, we pick the simplest function — a parabola.

The smoothing principle is illustrated in Fig. 7.5(a). Between coordinates $\bar{w}(\tau)$ and $\bar{u}$, the sharp bend of the specific line is replaced by a parabola. The smoothing parameter τ describes the offset from

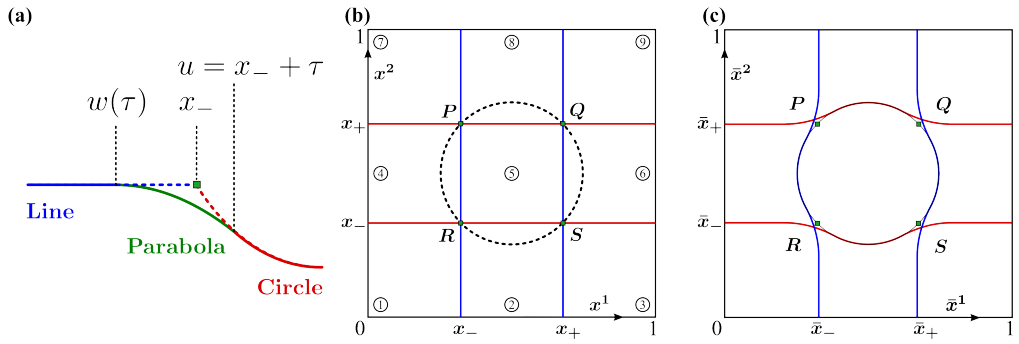


Figure 7.5.: *(a) Closeup of the "smoothed" transition from straight line to circle arc. Both parts are joined by an intermediate parabola piece. The parameter τ determines the junctures of the line elements — u depends directly on τ, and w follows indirectly — and is, thus, the measure for the transition's "smoothness". (b) Specific lines and characteristic points in transformed space, and (c) in Cartesian space for a smoothed mesh of a circular structure. Reproduced from [100].*

where parabola and circle arc meet and directly determines $\bar{u} = x_- + \tau$ (here, $x_\pm = 0.5 \pm r/\sqrt{2}$). Similar to the circle arc, the smoothed part of the specific line is described by parabola functions

$$\begin{aligned} PA_{LT/LB}(x^1,\tau) &= \pm m(\tau)\cdot(x^1 - w(\tau))^2 + x_\pm, \\ PA_{RT/RB}(x^1,\tau) &= \pm m(\tau)\cdot(x^1 - (1-w(\tau)))^2 + x_\pm, \end{aligned} \tag{7.15}$$

for the x^1 specific lines in Cartesian space and similar expressions for the x^2 specific lines except that PA_{LT} and PA_{RB} switch their expressions. Demanding continuity and differentiability at $\bar{u}$ (or $1-\bar{u}$) determines parabola slope m and coordinate $\bar{w}$ [100]. Picking the vertical x_- line as example, the corresponding specific line $\bar{x}_-$ in Cartesian space (for fixed τ) is given by

$$\bar{x}_- : \quad \bar{x}^2(x_-,x^2) = x^2, \qquad \bar{x}^1(x_-,x^2) = \begin{cases} x_-, & x^2 \in [0,w] \text{ or } [1-w,1], \\ PA_{LB}(x^2), & x^2 \in [w,u], \\ CA_L(x^2), & x^2 \in [u,1-u], \\ PA_{LT}(x^2), & x^2 \in [1-u,1-w]. \end{cases} \tag{7.16}$$

This smoothing is applied to all specific lines and all junctures. The resulting lines throughout the unit cell are depicted in Fig. 7.5(b). The actual coordinate transformation can be obtained with the help of the linear transition function LT as in the previous case. Therefore, the unit cell is divided into the regions depicted in Fig. 7.6, where the 15 considered subdomains for writing down the transformation functions are sketched. The regions for coordinate transformation $\bar{x}^1(x^1,x^2)$ are shown in panel (a), whereas the regions important for $\bar{x}^2(x^1,x^2)$ are shown in panel (b). It is important to notice that they differ. The corresponding detailed coordinate transformation functions can be found in Ref. [100].

As we will see in more detail later on, the smoothed mesh does not cause diverging effective material parameters anymore. However, it introduces additional discontinuities. These discontinuities are present in the non-differentiable mesh as well, but coincide there with the discontinuity of $\bar{\varepsilon}$. In

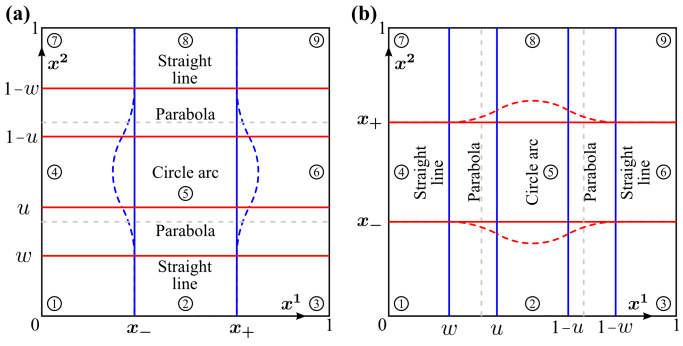


Figure 7.6.: *Sketch of the unit cell division into regions. (a) Partitioning for coordinate mapping $\bar{x}^1(x^1,x^2)$. (b) Partitioning for coordinate mapping $\bar{x}^2(x^1,x^2)$. The mapping principle is equivalent to the non-differentiable case, except for the altered specific lines with smooth parabola transitions.*

the non-differentiable case, they are, therefore, neither distinguishable nor bothering because they just slightly modify the already present discontinuity in the permittivity. However, they influence the previously not shown effective permeability in the same way. These discontinuities can be attributed to different coordinate line densities within and outside the circular structure. This difference occurs because the coordinate lines inside the structure are stretched by the linear transition, whereas the lines outside are compressed (cf. Fig. 7.2(b)). For instance, we consider the derivatives $\frac{\partial}{\partial x^2}\bar{x}^2(x^1, x^2)$ in regions ② and ⑤. Such derivatives constitute the basic summands of the effective material distributions in Eqs. (7.3). In region ② the derivative is given by

$$\frac{\partial \bar{x}^2}{\partial x^2} \overset{Eq.\ (7.13b)}{=} \frac{CA_B(x^1)}{x_-} \overset{x^1=0.5}{=} \frac{1-2r}{1-\sqrt{2}r}\,, \tag{7.17a}$$

and in region ⑤ the derivative reads

$$\frac{\partial \bar{x}^2}{\partial x^2} \overset{Eq.\ (7.14b)}{=} \frac{2\sqrt{r^2-(x^1-0.5)^2}}{(x_+ - x_-)} \overset{x^1=0.5}{=} \sqrt{2}\,. \tag{7.17b}$$

First of all, we conclude that the line density in both regions is independent of x^2 but varies along the structure surface (it depends on x^1). Evaluated at the exemplary point $x^1 = 0.5$ where the central coordinate line intersects the specific line, the discontinuity becomes apparent in the above equations.

Additional discontinuities are cumbersome to expand in a Fourier series and naturally introduce additional inaccuracies. This motivates a further alteration of the construction scheme.

Differentiable Meshes

The compression and stretching of the coordinate lines inside and outside the structure gives rise to discontinuous derivatives at the specific lines (structure surface). The core idea of differentiable meshes is the introduction of intermediate regions where the coordinate line density varies smoothly and is chosen such that it matches the densities of the surrounding regions. The new unit cell division is sketched in Fig. 7.7. The mapping in regions ① to ⑨ stays the same as in the non-differentiable (and smoothed) mesh construction schemes. However, even though we use the x_- and x_+ specific lines for the construction, we execute the mapping only in the smaller regions. For instance, in the case of region ② (cf. Eqs. (7.13)) we evaluate the mapping only for $x^1 \in [u, 1-u]$ and $x^2 \in [0, w]$ (instead of $x^2 \in [0, x_-]$).

The introduced intermediate regions are mapped according to new specific lines which can be derived from the specific lines of the smoothed mesh. If, for example, the smoothed specific line $\bar{x}_-$ is given by Eq. (7.16), then the derived new vertical specific lines are

$$\bar{w}: \quad \bar{x}^1(w, x^2) = LT(0, 0, x_-, \bar{x}_-, w), \qquad \bar{x}^2(w, x^2) = x^2, \tag{7.18a}$$

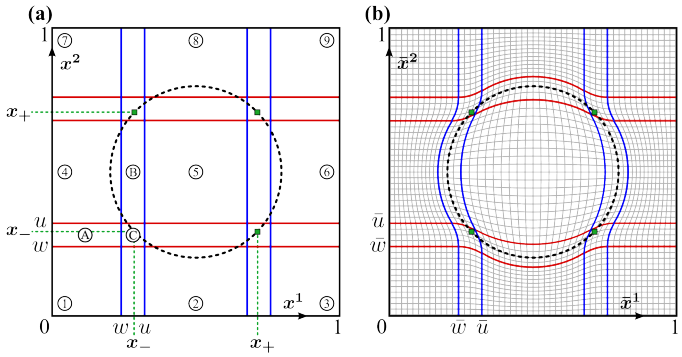


Figure 7.7.: *Specific lines and characteristic points in transformed space and in Cartesian space for the construction of a differentiable mesh of a circular structure. The specific lines are constructed similar as for the smoothed meshes. Instead of a single specific line covering the structure's surface, every surface section is handled by two specific lines, one on each side of the surface.*

and

$$\bar{u}: \quad \bar{x}^1(u, x^2) = LT(x_-, \bar{x}_-, x_+, \bar{x}_+, x^1)\Big|_{x^1=u},$$

$$\bar{x}^2(u, x^2) = \begin{cases} LT(0, 0, w, \bar{w}(x^1), x^2)\Big|_{x^1=u}, & x^2 \in [0, w] \text{ or } [1-w, 1], \\ C(x^1, x^2)\Big|_{x^1=u}, & x^2 \in [w, u], \\ LT(x_-, \bar{x}_-, x_+, \bar{x}_+, x^2)\Big|_{x^1=u}, & x^2 \in [u, 1-u], \\ C(x^1, 1-x^2)\Big|_{x^1=u}, & x^2 \in [1-u, 1-w], \end{cases} \tag{7.18b}$$

where $C(x^1, x^2)$ is a function derived from the yet unknown mapping in region ℂ. The specific lines x_-, $\bar{x}_-$, x_+, and $\bar{x}_+$ are functions of one transformed coordinate — which one depends on the considered Cartesian coordinate: if the $\bar{x}^1$ coordinate is considered they depend on x^2, or vice versa.

For symmetry reasons, we can limit the discussion to regions Ⓐ to ℂ. In region Ⓐ the mapping is then given by

$$\bar{x}^1(x^1, x^2) = LT(0, 0, w, \bar{w}(x^2), x^1), \qquad \bar{x}^2(x^1, x^2) = x^2, \qquad (x^1, x^2) \in \text{Ⓐ}\,. \tag{7.19}$$

The mapping of coordinate x^2 in region Ⓑ is straightforward as well. We can write it down as a linear transition between the parabolic specific lines $\bar{u}$ and $1-\bar{u}$, which ensures a continuous x^2 coordinate line density variation with respect to x^1 between regions ④ and ⑤ by construction:

$$\bar{x}^2(x^1, x^2) = LT(u, \bar{u}(x^1), 1-u, 1-\bar{u}(x^1), x^2), \qquad (x^1, x^2) \in \text{Ⓑ}\,. \tag{7.20}$$

However, the x^1 coordinate mapping is more complicated. We recall that the line density ($\frac{\partial}{\partial x^1}\bar{x}^1$) in ④ and ⑤ was different. Furthermore, both densities varied with x^2 independently. A simple LT would provide again a third distinct density value constant in x^1. Instead, in order to get rid of discontinuous derivatives, it must be ensured that the density changes continuously from left to right

and bottom to top (the latter is guaranteed by Eq. (7.20)). The related boundary conditions that have to be met by the mapping can be summarized as follows. Continuity requires

$$\bar{x}^1(w, x^2) \overset{!}{=} \bar{x}^1(w, x^2)\Big|_{④} = \bar{w}(x^2) \,, \qquad \bar{x}^1(u, x^2) \overset{!}{=} \bar{x}^1(u, x^2)\Big|_{④} = \bar{u}(x^2) \,, \tag{7.21a}$$

whereas differentiabilty requires

$$\left.\frac{\partial \bar{x}^1}{\partial x^1}\right|_{(w,x^2)} \overset{!}{=} \left.\frac{\partial \bar{x}^1}{\partial x^1}\right|_{(w,x^2),④} \,, \qquad \left.\frac{\partial \bar{x}^1}{\partial x^1}\right|_{(w,x^2)} \overset{!}{=} \left.\frac{\partial \bar{x}^1}{\partial x^1}\right|_{(u,x^2),⑤} \,, \tag{7.21b}$$

$$\left.\frac{\partial \bar{x}^1}{\partial x^2}\right|_{(w,x^2)} \overset{!}{=} \left.\frac{\partial \bar{x}^1}{\partial x^2}\right|_{(w,x^2),④} \,, \qquad \left.\frac{\partial \bar{x}^1}{\partial x^2}\right|_{(w,x^2)} \overset{!}{=} \left.\frac{\partial \bar{x}^1}{\partial x^2}\right|_{(u,x^2),⑤} \,. \tag{7.21c}$$

Both $\bar{w}$ and $\bar{u}$ describe ellipse arcs in the considered regimes.

For region Ⓒ boundary conditions similar to Eqs. (7.21) can be easily derived substituting ④ with Ⓐ and ⑤ with Ⓑ. By virtue of symmetry in the coordinates, it is sufficient to set up the boundary conditions for one coordinate only, e.g., $\bar{x}^1$ with left boundary provided by the specific line $\bar{w}$ and right boundary from the specific line $\bar{u}$. Any smooth mapping function fulfilling these conditions is suitable to provide a fully differentiable mesh.

Küchenmeister [100] provided an elegant ansatz to solve this problem. For details, we refer the interested reader to the paper.

With the differentiable mesh we have completed the introduction of analytic AC transformations. We defined mapping functions throughout the entire unit cell which map any coordinate between transformed and Cartesian space. The second ingredient to adapted meshes and the missing piece is the adaptive spatial resolution which is introduced in the subsequent section.

Adaptive Spatial Resolution

The adaptive spatial resolution (ASR) concept describes the concerted local enhancement of sampling points for the discrete representation of the physical structure. In particular, this increase in the density of coordinate lines is interesting in the vicinity of material interfaces where the permittivity displays jump discontinuities and the finite Fourier series converges poorly. The observation is that a local increase in resolution eventually improves this convergence behavior [97]. For instance, this improvement could be accomplished by an effective structure in transformed space whose higher order coefficients decay faster than those of the original structure.[12]

Adaptive spatial resolution is implemented by additional one-dimensional coordinate transformations, which are independently applied to the two coordinates x^1 and x^2 before the AC mapping.

[12] In the ideal case, the effective structure could be described bandwidth limited functions. The achievement of a bandwidth limited — essentially entirely smooth — permittivity might be possible by construction of concurrent jumps with a suitably designed coordinate transformation. However, it is reasonable that the original smooth permeability function then picks up a jump discontinuity. This is due to the fact that the same spatial transformation always applies to both permittivity and permeability at the same time. From this perspective, a bandwidth limited Fourier representation of the problem by a properly adapted coordinate transformation seems unrealistic. At most, some of the discontinuity might be shifted from permittivity to permeability such that the entire problem converges faster. Besides these general considerations, the ASR we present in the course of this thesis is in no way designed to achieve this ultimate goal. Anyway, an ASR transformation alone is unlikely to be sufficient for this purpose.

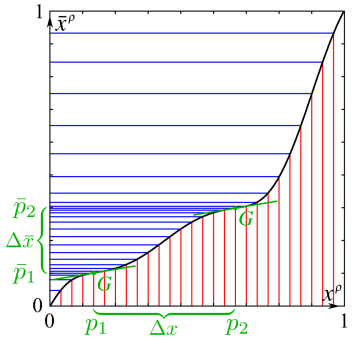


Figure 7.8.: *Sketch of the principle of adaptive spatial resolution (ASR). We define a one dimensional smooth mapping function (black curve) through points $(x_1, \bar{x}_1)$ and $(x_2, \bar{x}_2)$ with slope G at those points ($G < 1$). The function maps equidistantly spaced coordinates (red vertical lines) to spatially adapted non-equidistant new coordinates (blue horizontal lines). Thus, the coordinate density at $\bar{x}_1$ and $\bar{x}_2$ is locally increased. The ASR strength depends on the slope G.*

These ASR transformation functions map the equidistantly spaced sampling points x_k^ρ, with $k = 0, 1, \ldots, N_{\text{fft}} - 1$ required by the FFT (cf. Eqs. (3.21)) to spatially adapted, non-equidistant new sampling coordinates $\bar{x}_k^\rho = \bar{x}^\rho(x_k^\rho)$. This ASR principle is illustrated in Fig. 7.8, where red vertical lines highlight the equidistant sampling positions x_k^ρ and horizontal blue lines the $\bar{x}_k^\rho$. Please note that the calculated new sampling points $\bar{x}_k^\rho$ adopt the role of the unbarred coordinates in the subsequent discretization of the AC mapping.

The compression points on the vertical axis $\bar{p}_1$ and $\bar{p}_2$ are fixed by the positions of the structure surface at which we want to increase the coordinate line density. These surfaces are given by of the specific lines of the AC mapping, for example. The inflection point coordinates on the horizontal axis p_1 and p_2 can be chosen freely. Their position determines the structure surfaces and effective geometry in transformed space and, thereby, the fraction of coordinate lines within the three intervals $[0, \bar{p}_1]$, $[\bar{p}_1, \bar{p}_2]$, and $[\bar{p}_2, 1]$ on the vertical axis. The quantities $\Delta\bar{x}$ and Δx are of particular interest. The former represents the physical size of the object in Cartesian space, whereas the latter determines the object size in the transformed space. The coordinate line density in Cartesian space is inversely proportional to the slope of the transformation function. Consequently, the slope G at the inflection points is one of the parameters we are interested in.

A concrete proposal for such transformation functions has been made by Vallius [98]. This so-called *Vallius transformation* reads

$$\bar{x}(x) = \alpha + \beta x + \frac{\gamma}{2\pi} \sin\left(2\pi \frac{x - p_{l-1}}{p_l - p_{l-1}}\right), \qquad x \in [p_{l-1}, p_l], \quad l = 2, ..., n\,, \tag{7.22a}$$

with

$$\alpha = \frac{p_l \bar{p}_{l-1} - p_{l-1}\bar{p}_l}{p_l - p_{l-1}}, \qquad \beta = \frac{\bar{p}_l - \bar{p}_{l-1}}{p_l - p_{l-1}}, \qquad \gamma = \left(p_l - p_{l-1}\right) G - \left(\bar{p}_l - \bar{p}_{l-1}\right), \tag{7.22b}$$

where $\{(p_l, \bar{p}_l) : \quad l = 1, ..., n\}$ denote the coordinates of the inflection points. Equations (7.22) essentially describe a continuous, piecewise linear function through the inflection points which are

modulated with a sine. The single pieces are connected at the inflection points such that the entire transformation function is smooth.

In order to maintain the required periodicity with period L, Vallius chooses $(p_1, \bar{p}_1) = (0, 0)$ and $(p_2, \bar{p}_2) = (L, L)$ (cf. Fig. 7.9(a)). This means that the coordinate lines are always compressed at the unit cell edges. This, in turn, requires that the unit cell is chosen such that one material surface per direction coincides with the unit cell boundary. For the structures considered in Ref. [98] this is intended. We, on the other hand, would like to have more freedom in the choice of the unit cell.

Our approach to a *generalized Vallius transformation* is sketched in Fig. 7.9. The goal of the illustration example is the construction of a compression function for a structure with two material interfaces ($n = 2$) within the unit cell at $\bar{p}_1 = 0.1$ and $\bar{p}_2 = 0.3$ and structure size $\Delta\bar{x} = \bar{p}_2 - \bar{p}_1 = 0.2$. To fulfill the periodicity requirements, the final compression function must contain the boundary points $(0, 0)$ and(L, L) like before. We start with the original Vallius transformation and blow up the structure in transformed space by a factor of 2.25 to the size $\Delta x = p_2 - p_1 = 0.45$. Furthermore, we arbitrarily pick a compression $G = 0.15$.

The first step is a vertical shift of the entire function to the final positions of the inflection points on the vertical axis (cf. Fig. 7.9(b)). The intersection point $\tilde{x}$ with the boundary $\bar{x} = L$ must be computed numerically because the evaluated equation is transcendental. In the second step, depicted in Fig. 7.9(c), the entire function is shifted to the right by $\tilde{L} = L - \tilde{x}$. Because of periodic boundary conditions, the red part of the function outside the interval is finally shifted back by one period into the considered region. The result is depicted in Fig. 7.9(d). For more details and formulas about the generalized Vallius transformation we refer the interested reader to Ref. [100]. With this approach we can construct Vallius-like compression functions with an arbitrary number of inflection points at any position within the unit cell. We would like to note that it is important to pick the coordinates of subsequent inflection points in a monotonically increasing order, and furthermore pay attention to a

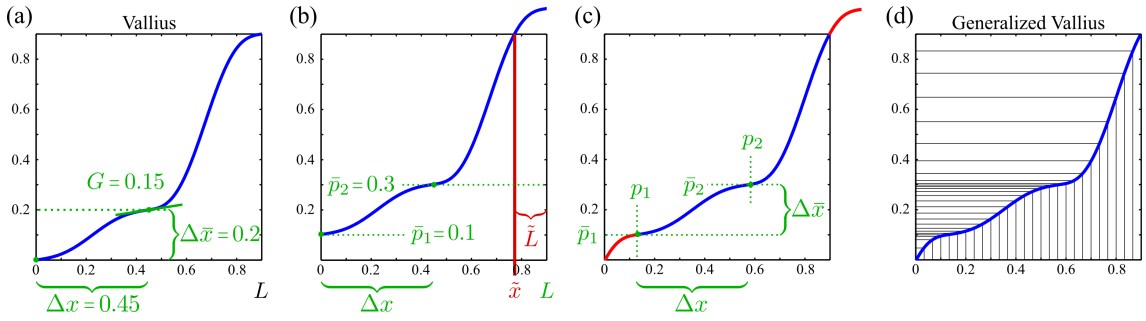


Figure 7.9.: *Construction of the generalized Vallius transformation illustrated with an example. In order to have the ASR compression inside the unit cell at the specific line locations of the adapted mesh, the Vallius transformation function has to be shifted. (a) The entire original function is first moved "up" to the desired specific line coordinates $\bar{p}_1 = 0.1$ and $\bar{p}_2 = 0.3$ depicted in (b). Subsequently, the intersection point $\tilde{x}$ of the function with the unit cell edge is (numerically) determined. (c) The entire function is moved "right" by $\tilde{L} = L - \tilde{x}$. Because of periodic boundary conditions, the red part of the function outside the interval (upper right) is shifted back by one lattice constant into the interval (lower left). (d) Final generalized Vallius transformation function. Picture adapted from Ref. [100].*

monotonically growing transformation function in general by proper choice of G, in order to avoid folding of the adapted mesh.

With the construction principle of ASR we conclude the construction of analytically generated adapted meshes. In the next section we focus on their influence on the effective material parameters in the transformed space and give the reader an impression what the effective material parameters look like.

Comparison of Analytic Meshes

We presented three different but related approaches for the analytic construction of adapted meshes above. In the motivation for the smoothed and and differentiable construction schemes we already mentioned discontinuities and other obstacles. This section is intended to illustrate the differences by means of the material distributions in transformed space, and carve out the positive and negative aspects of each approach.

We discuss these issues with the help of the already introduced circular structure (cf. Fig. 7.2(a)). Figure 7.10 depicts the three mesh types (top row), the ε^{11} components (center row), and the μ^{11} components (bottom row). The non-differentiable mesh in column (a) resolves the material surface without staircasing effects. Thus, the effective permittivity exhibits a perfectly grid-aligned square structure. As already mentioned, the permittivity values in the corners of the square diverge. In the plot the colorbar is saturated at five in order to highlight the structural details. The corresponding permeability tensor component underneath shows a very similar behavior. The only difference is the scaling factor of one half within the quadratic region because the original permittivity in Cartesian space is unity whereas the original permittivity in that region is two.

In comparison, for the smoothed mesh in column (b) we ensured by the parabolic smoothing that the coordinate lines are differentiable. With the smoothing we give up an exact representation of the structure surface. Hence, the material interface has now the shape of a rounded square and a small dose of staircasing is unavoidable. As a direct consequence from differentiable coordinate lines, mixed derivatives, e.g., $\frac{\partial \bar{x}^1}{\partial x^2}$, become continuous, whereas alike derivatives, as for example $\frac{\partial \bar{x}^1}{\partial x^1}$, which can be considered as coordinate line densities, still have jump discontinuities. This mismatch is visible in the smoothed mesh when going from left to right along the horizontal central coordinate line: the densities of vertical lines between outside the circular structure and inside are different at the material surface. This comes from the fact that outside the structure region the LT function compresses the lines, whereas inside it stretches them compared to the equidistant Cartesian mesh. Since both alike and mixed derivative terms are summed up in every effective material tensor component (cf. Eqs. (7.3)), this density difference becomes visible as square shadow structure introducing additional jump discontinuities in permittivity and permeability. It is particularly well visible in the permeability, where the much larger discontinuity of the physical structure is absent. The coordinate line smoothing thus spatially separates the discontinuities stemming from the different materials and those from the coordinate line density mismatch which coincide in the case of the non-differentiable mesh. However, due to the smoothing, the coordinate lines in the corners are no longer parallel, the Jacobian determinant does not vanish anymore, and the maximum effective permittivity is about twice the original value.

Last but not least we discuss the differentiable mesh in column (c). As we can see, the effort of constructing a mesh with a continuous variation of the coordinate line density was worth the trouble. The

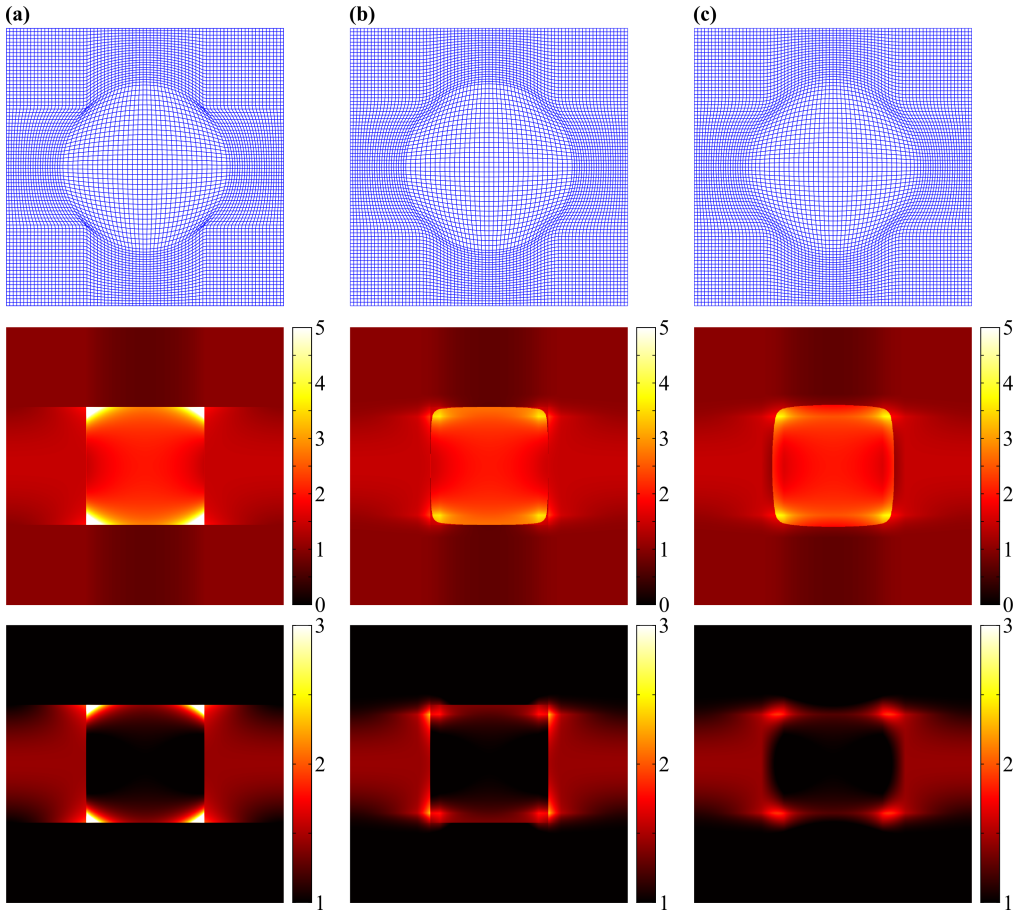


Figure 7.10.: *Comparison of different mesh types and the resulting material tensor distributions. (a) Non-differentiable mesh, (b) smoothed mesh, and (c) differentiable mesh. The first row depicts the meshes ($\tau = 0.035$), the second row shows the effective permittivity ε^{11}, and the effective permeability μ^{11} is depicted in the last row. Effective permeability and permittivity plots show 1024×1024 sampling points. Note the different colorbars.*

effective permeability distribution clearly shows no discontinuities anymore and the square shadow structure in the permittivity is gone as well. Besides that improvement, the fully differentiable mesh exhibits the same qualitative behavior as the smoothed mesh, even though the material interface of the structure is slightly less grid-aligned that the non-differentiable mesh and, therefore, shows a bit more staircasing.

In practice, the three different techniques reduce to essentially two: the non-differentiable mesh and the differentiable mesh. The smooth mesh can be seen as an intermediate step in the construction which inherits neither the grid-aligned exact surface representation of the non-differentiable mesh, nor the fully differentiable mapping of the differentiable mesh. Hence, it rather complicates the Fourier representation by additional jump discontinuities. The performance of the two remaining techniques in real applications will be one of the topics covered in Chap. 8.

7.1.2. Automated Adapted Mesh Generation

The analytical mesh construction is not the only way to obtain adapted meshes. In particular, for complicated geometries it might actually be rather tedious. An illustrative example is the cross section of a photonic crystal fiber, where the silicon matrix is interspersed with air holes arranged in a hexagonal lattice. Such fibers usually comprise several rings of air holes — roughly about seven — which already makes a total of 148 single air holes provided the central hole is left out. This level of complexity is the natural domain of automated adapted mesh generation as introduced by Essig *et al.* [104]. Since we do not use the technique in the course of this work, we only briefly sketch the procedure.

The core element of automated mesh generation is a potential energy landscape throughout the unit cell with local minima at material surfaces. In a simple picture, the mesh is then modeled as spring-mass system in the landscape, where each sampling point (intersection of coordinate lines) is represented by a mass, and the connection to nearest neighbors is represented by a spring. The masses in the landscape provide for the adaption of the mesh to the structure, whereas the springs act as restoring force to prevent that all masses accumulate in the minima of the landscape. Starting from a regular mesh, the position of the sampling points will then evolve in a way as to minimize the total energy of the system, which is given by the mechanical energy stored in the total ensemble of springs and the summarized gravitational energy of the masses.

While this picture gives a good illustration of the general idea, it turns out that it is too simple to produce nice coordinate transformations. Thus, in reality the meshing relies on the minimization of a fictitious energy functional with four distinct contributions [104]

$$\mathcal{E}\Big(\bar{x}^1(x^1,x^2),\bar{x}^2(x^1,x^2)\Big) = \int_{\mathrm{UC}} \mathrm{d}x^1\,\mathrm{d}x^2\,\Big(\mathcal{E}_c(x^1,x^2)+\mathcal{E}_s(x^1,x^2)+\mathcal{E}_g(x^1,x^2)+\mathcal{E}_t(x^1,x^2)\Big)\,. \tag{7.23}$$

The *compression energy* term

$$\mathcal{E}_c(x^1,x^2) = s_c\cdot\det(g^{\rho\sigma}) \tag{7.24}$$

has a similar role like the springs in the simple picture — it exerts a restoring force towards the original regular mesh in the direction along the coordinate line between two adjacent mesh points. The restoring force perpendicular to this coordinate line is ensured by the *shear energy* term

$$\mathcal{E}_s(x^1,x^2) = s_s\cdot\mathrm{tr}(g^{\rho\sigma})\,. \tag{7.25}$$

The remaining two terms, *gradient energy*

$$\mathcal{E}_g(x^1,x^2) = -\Big|\bar{\boldsymbol{\nabla}}\cdot S_{sm}\Big(\bar{x}^1(x^1,x^2),\bar{x}^2(x^1,x^2)\Big)\Big|\,, \tag{7.26}$$

and *tangential energy*

$$\mathcal{E}_t(x^1,x^2) = s_t\cdot\left(\Big|(\bar{\boldsymbol{\nabla}}\cdot S_{sm})\cdot\mathbf{e}_1\Big|^2+\Big|(\bar{\boldsymbol{\nabla}}\cdot S_{sm})\cdot\mathbf{e}_2\Big|^2\right)\,, \tag{7.27}$$

provide for the adaption of the mesh to the structure. The procedure, the resulting mesh and effective material distributions are illustrated in Fig. 7.11.[13] The gradient energy relies on the structure function S_{sm} which is essentially the normalized permittivity distribution with a Gauss'ian smoothing

[13] Pictures depicted in Fig. 7.11 (a), (b), and (c) adapted from Ref. [65] by courtesy of Sabine Essig.

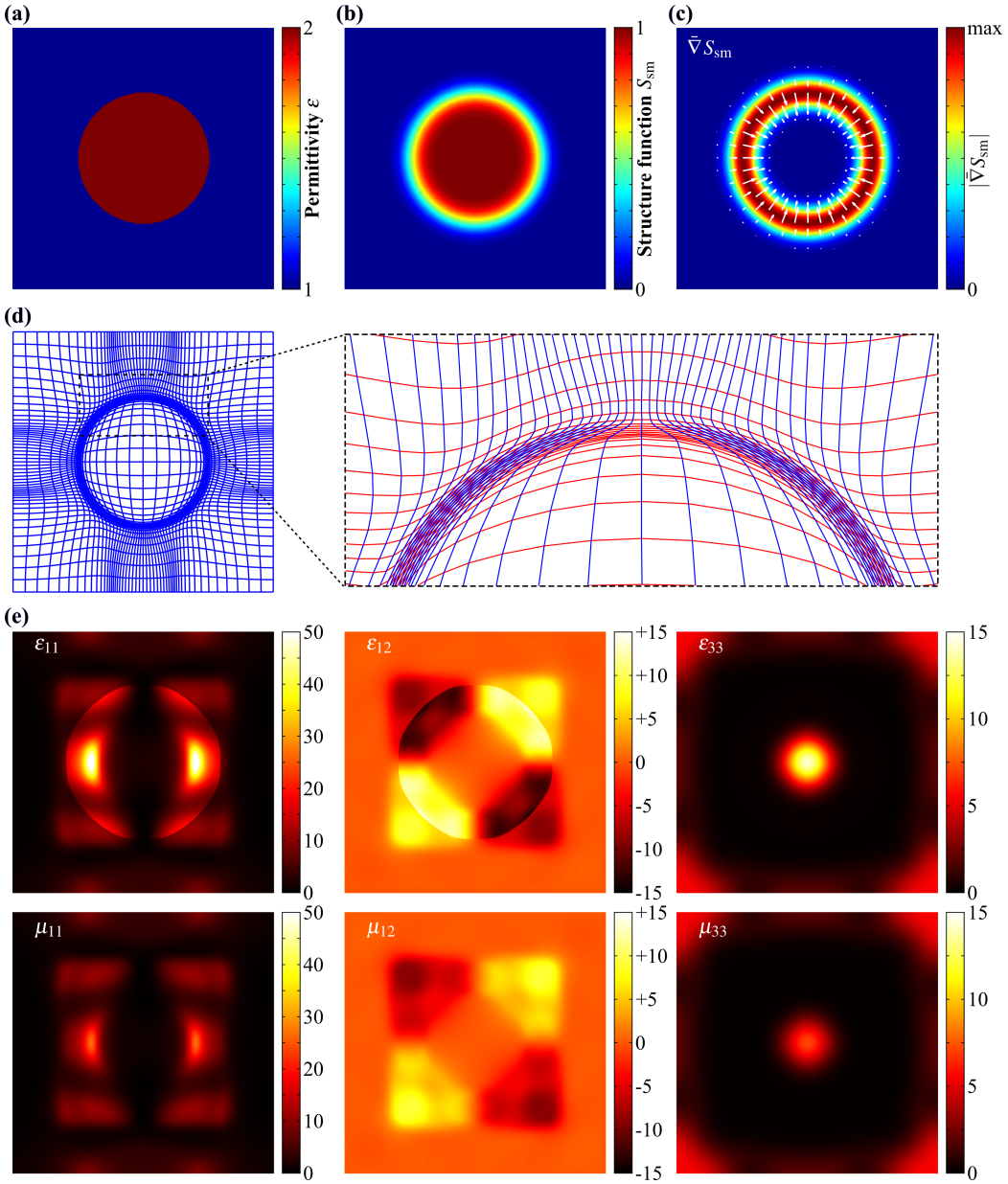


Figure 7.11.: *Automated adaptive mesh generation. Panel (a) depicts the permittivity in Cartesian space. The structure function S_{sm}, shown in panel (b), is derived from the permittivity with a Gaussian smoothing of the material interfaces. Panel (c) illustrates its gradient field with little white arrows and the magnitude of the gradient field color coded. The resulting adapted mesh with the minimized energy functional $\mathcal{E}$ is plotted in (d). Horizontal coordinate lines in the zoom on the right hand side are highlighted in red to improve the distinguishability from vertical lines. (e) Selection of related material distributions throughout the unit cell in transformed space. In contrast to the analytically generated non-differentiable mesh, the minimized mesh does not provide grid aligned structures (cf. Fig.* 7.10*).*

of the discontinuous jumps at the surface (cf. Fig. 7.11(b)). Hence, the gradient of the smoothed structure function is zero in homogeneous regions and non-zero near material surfaces with its maximum absolute value exactly at the surface (cf. Fig. 7.11(c)). With the minus sign in Eq. (7.26) one achieves that the mesh lines tend to accumulate at the surface because of a favorable energy constellation. In contrast, the tangential energy term is a penalty term which punishes coordinate lines — represented by their covariant basis vectors $\mathbf{e}_\rho$ — which are *not* orthogonal to the gradient field. Since the gradient is always perpendicular to the structure surface, the coordinate lines, thus, prefer the desired parallel surface alignment (cf. inset of Fig. 7.11(d)). The strength parameters s_c, s_s, and s_t allow for an adjustment of the respective contributions relative to the gradient energy. In practical applications these parameters have to be properly balanced in order to achieve nice meshes.

The automatically adapted mesh generation produces inherently smooth meshes and automatically incorporates an enhanced adaptive spatial resolution at material interfaces (cf. Sec. 7.1.1). Such a mesh is depicted in Fig. 7.11(d). The current state of art is a normalized smoothed structure function S_{sm} which is represented in Fourier space and whose width of the smoothing region can be adjusted by a parameter [104]. Future developments could focus on an improved potential landscape design which respects different permittivity values and more sophisticated gradient fields such that gratuitous mesh deformation is reduced which sometimes leads to high values in effective permeability and effective permittivity. As can be seen from Fig. 7.11(e), the automated mesh does not necessarily produce grid-aligned effective structures.

7.2. Stretched Coordinate Perfectly Matched Layers

The Fourier modal method naturally incorporates (Bloch-) periodic boundary conditions due to the plane wave expansion and the use of Fourier transformations. This makes the method predestined for the simulation of periodic structures like, for example, photonic crystals. Furthermore, since the set of basis functions remains the same in every layer, the matching in the scattering matrix algorithm is very robust and simple. However, this scattering matrix algorithm seems often undervalued. Besides the modal expansion which allows for efficient simulation of x^3-invariant structures, a particularly outstanding strength of the method is in fact this s-matrix algorithm. It allows for a very efficient simulation of repeated patterns along the propagation direction by recycling of the respective s-matrices. If we look at the FMM in such light, it is desirable to use the method for systems with simple cross sections but large aspect ratios as well.

The required (artificial) periodic boundary conditions in the lateral plane reduce the FMM's application scenarios to aperiodic systems with no or very small scattering processes. Such scenarios, for instance, can often be found in (wave-) guiding based applications, e.g., the determination of guided eigenmodes which evanescently decay in lateral directions, or mode couplers where forward or backward scattering within the guiding structure exceeds outward scattering by far. For such systems, an artificial periodicity is introduced and the structures are set sufficiently apart. The procedure is schematically depicted in Fig. 7.12. Other systems where strong scattering plays a non-negligible role cannot be simulated with the ordinary periodic FMM. In these systems the cross talk and energy exchange between neighboring unit cells leads to strong coupling effects which usually significantly influence the overall result.

It is known that Finite Element like methods are commonly more efficient in the determination of

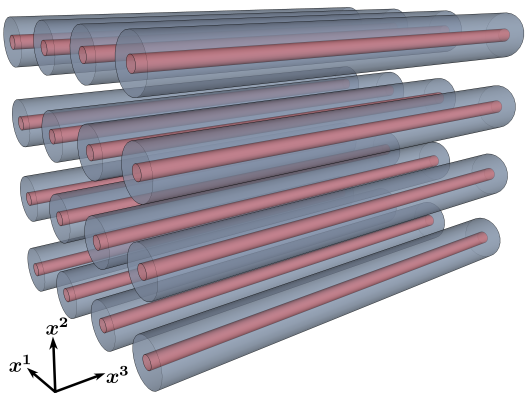


Figure 7.12.: *Sketch of the artificial periodic arrangement of a waveguide structure in the FMM.*

eigenmodes of aperiodic structures than the FMM [74]. They lack the periodicity requirement of the FMM, and absorbing boundary conditions to mimic *open boundaries* are well established for polynomial bases [105]. So, the eligible question arises why the FMM, which is a reliable and established simulation tool for periodic structures, should be exploited in this direction, if an FEM eigenmode solver can be combined with a scattering-matrix code as well. Surprisingly, it turns out that the layer matching by basis functions (cf. Sec. 5.3.1) inherent to the FMM s-matrix algorithm is more accurate than real-space matching schemes required by FEM methods [74], which is the limiting factor in transmission calculations — here tested with the B-spline modal method.[14] This is the case even though the individual eigenmodes calculated by the BMM converge faster by several orders.

The lifting of the FMM application restriction to aperiodic systems with weak scattering can be achieved by the incorporation of open boundary conditions. What might sound easy is in fact a challenging task since, in order to keep the FMM concept, finite unit cells and periodic boundary conditions must be maintained. This squaring the circle problem requires a "trick" — *stretched coordinate transformations*.

The core idea of stretched-coordinate transformations is the compression of infinite space onto a finite edge layer. The principle is illustrated in Fig. 7.13. Here, we use the same technique as for the adapted meshes before. The main difference is that instead of a mapping of a finite interval onto itself, we map the infinite space onto a finite interval of size d or vice versa:

$$\left[-\infty,+\infty\right] \quad \longleftrightarrow \quad \left[-\frac{d}{2},+\frac{d}{2}\right] . \tag{7.28}$$

To this end, we define a mapping function

$$F_\rho(x^\rho): \qquad \left[-\frac{d}{2},+\frac{d}{2}\right] \quad \longmapsto \quad \left[-\infty,+\infty\right] . \tag{7.29}$$

The stretched-coordinate mapping excludes a central domain of size e ($F(x^\rho) = x^\rho$) where the mapping is just the identity relation and all physical properties of the problem we are interested

[14] FEM methods require real-space matching schemes in the sense that they intentionally use grids which are specifically adapted to the structures in the layer. Thus, the basis functions usually differ in adjacent layers.

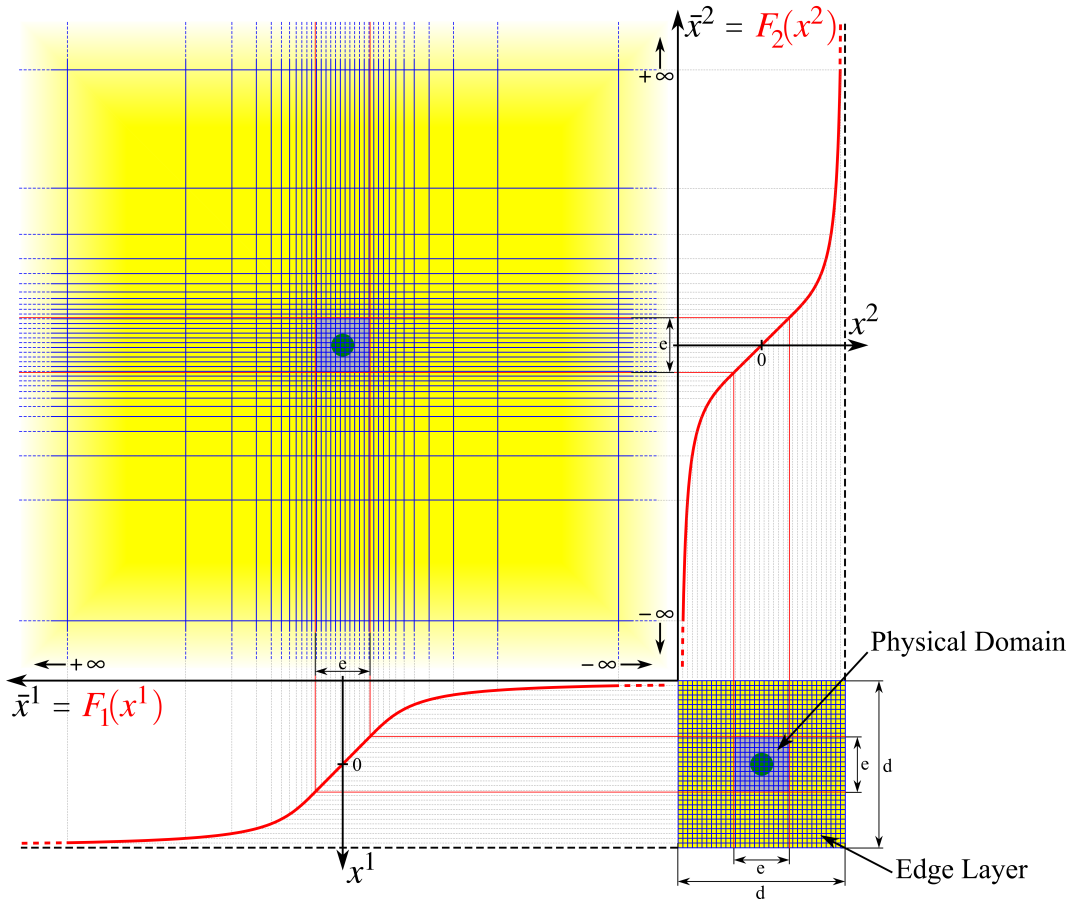


Figure 7.13.: *Two-dimensional stretched-coordinate transformation. The infinite domain (top left) is mapped onto a finite unit cell of size d (bottom right) — or vice versa. The physical domain of size e (grey shaded) remains untransformed (identity mapping), the surrounding infinite space is squeezed into the finite edge layer (yellow shaded). Considering it the other way round, the equidistant mesh (blue lines) in transformed space is mapped onto a non-equidistant grid with decreasing coordinate line density with increasing distance from the physical domain.*

in are evaluated. Moving from this inner domain outwards, the coordinate compression gradually increases, finally reaching infinity at the outer boundary.

The calculated solutions within the edge layer, on the other hand, are interesting only as they should mimic the solutions within the infinite free space to a certain degree of accuracy. In the ideal case, radiated energy traveling into the edge layer should neither be reflected at the interface nor ever return or disturb adjacent unit cells. Unfortunately, this is not the case as long as the coordinate mapping is real. First, in contrast to time domain, in frequency domain the radiative modes always (instantly) reach to infinity. Second, the compression of the infinite space also compresses the radiative solutions. Thus, spatial frequencies occur way beyond the limit of what can be correctly represented by reasonable mesh resolutions and the truncation order of feasible Fourier series. While Shyroki [106] claims this ansatz works in time-domain methods, we believe it cannot properly work

in frequency-domain methods because stationary solutions reach to infinity and are influenced by the issues mentioned above.

The crucial step to overcome these obstacles is the adding of absorption. If the amplitude of radiative solutions decays "faster" than the increase in spatial frequency in the squeezed space, the errors due to limited resolution and a truncated Fourier series decrease. Absorbing boundary conditions, where the edge layer consists of an artificial absorbing material instead of free space, are the most basic approach. Berenger [107] found out that spurious reflections, in particular for oblique incidence onto the surface between the physical region and the absorptive edge layer, can be eliminated if the materials are impedance matched.[15] With this discovery he coined the technique of *perfectly matched layers* (PML). Berenger, however, did not employ coordinate transformations for his PMLs but only artificial absorptive materials.

The introduction of artificial absorptive materials with complex material parameters leads to wave vectors $\mathbf{k}$ with complex amplitudes $k = k_\mathrm{R} + \mathrm{i}k_\mathrm{I}$, $k_\mathrm{R}, k_\mathrm{I} > 0$ (cf. Eq. (2.30)). For the exponential term of a forward propagating one dimensional plane wave,

$$e^{+\mathrm{i}kx} \quad \longrightarrow \quad e^{+\mathrm{i}k_\mathrm{R}x} \cdot e^{-k_\mathrm{I}x}\,, \tag{7.30}$$

this implicates the desired exponential decay and electromagnetically isolates the unit cell from its neighbors. However, a similar effect can easily be achieved if — instead of an absorptive material — the edge layer is constructed via a complex-valued coordinate transformation. The exponential term from Eq. (7.30) with a complex coordinate $x = x_\mathrm{R} + \mathrm{i}x_\mathrm{I}$ and a real wave vector $k = k_\mathrm{R}$

$$e^{+\mathrm{i}kx} \quad \longrightarrow \quad e^{+\mathrm{i}k_\mathrm{R}x_\mathrm{R}} \cdot e^{-k_\mathrm{R}x_\mathrm{I}} \tag{7.31}$$

gains an equivalent decay. Furthermore, since the physical material on both sides of the interface between edge layer and physical domain now remains the same, the created *stretched-coordinate perfectly matched layer* (SC-PML) is automatically impedance matched at its surface. It was Chew and Weedon [108] and Sacks *et al.* [109] who established the PML in a stretched-coordinate formulation — with very simple complex mappings at first. Using complex stretched-coordinates, the impedance matching is fulfilled for the entire SC-PML domain as well, because a coordinate transformation always transforms permittivity and permeability in the same way. Chew *et al.* [110] later demonstrated that SC-PMLs are reflectionless irrespective of the chosen coordinate transformation function, propagation direction, or polarization of the electromagnetic wave. Luckily, the SC-PML can be formulated such that it resorts to the same implementation developed for the adaptive meshes. The FMM equipped with PMLs is commonly referred to as the *aperiodic Fourier modal method* (aFMM).

The general SC-PML mapping formalism and the connection between different popular approaches on a formal level is sketched in Sec. 7.2.1. Thereafter, Sec. 7.2.2 introduces effective permittivity and effective permeability in the squeezed space for the covariant form of Maxwell's equations in analogy to Sec. 7.1. Two different specifications of stretched-coordinate transformation functions are outlined in the subsequent paragraphs. Section 7.2.3 addresses the original PML proposal for the FMM by Lalanne *et al.* [99], and Sec. 7.2.4 transfers the *complex frequency shifted perfectly matched layers* (CFS-PML) concept, a popular PML type predominantly used in time domain methods, to the realm of the Fourier modal method.

[15] The wave impedance for a dielectric material is defined as $Z = \sqrt{\frac{\mu(\mathbf{r},\omega)}{\varepsilon(\mathbf{r},\omega)}}$.

7.2.1. Coordinate Mapping

The *infinite* physical system we want to simulate is described by the coordinate system $\mathcal{O}\bar{x}^1\bar{x}^2\bar{x}^3$. In our case, this is a *subspace* of the infinite complex three dimensional space $\mathbb{C}^3$. This physical subspace is transformed onto a three dimensional *finite* domain in $\mathbb{R}^3$ described by the coordinate system $\mathcal{O}x^1x^2x^3$.

As proposed by Lalanne *et al.* [99], for the special case of separable and independent coordinate transformations in both dimensions of the transversal plane, the mapping can in general be noted down as

$$\bar{x}^1 = F_1(x^1)\,, \tag{7.32a}$$
$$\bar{x}^2 = F_2(x^2)\,, \tag{7.32b}$$
$$\bar{x}^3 = x^3\,, \tag{7.32c}$$

where the complex valued functions F_ρ are called the coordinate stretching functions. From Eq. (7.6) we deduce that in this case the derivatives with respect to the original coordinates $\bar{x}^\rho$,

$$\frac{\partial x^1}{\partial \bar{x}^1} = \frac{\frac{\partial \bar{x}^2}{\partial x^2}}{\frac{\partial \bar{x}^1}{\partial x^1}\frac{\partial \bar{x}^2}{\partial x^2}} = \left(\frac{\partial \bar{x}^1}{\partial x^1}\right)^{-1}\,, \tag{7.33a}$$
$$\frac{\partial x^2}{\partial \bar{x}^2} = \frac{\frac{\partial \bar{x}^1}{\partial x^1}}{\frac{\partial \bar{x}^1}{\partial x^1}\frac{\partial \bar{x}^2}{\partial x^2}} = \left(\frac{\partial \bar{x}^2}{\partial x^2}\right)^{-1}\,, \tag{7.33b}$$

can be replaced by the inverse of the derivative with respect to the transformed coordinates x^ρ. This means we can conveniently use the inverse of the derivatives of Eqs. (7.32):

$$\frac{\partial x^\rho}{\partial \bar{x}^\rho} \overset{\text{Eqs. (7.33)}}{=} \left(\frac{\partial \bar{x}^\rho}{\partial x^\rho}\right)^{-1} \overset{\text{Eqs. (7.32)}}{=} \left(\frac{\partial F_\rho}{\partial x^\rho}\right)^{-1} \equiv f_\rho(x^\rho)\,. \tag{7.34}$$

Due to the independent one-dimensional transformations all mixed derivatives vanish.

Teixeira *et al.* [105, 111] describe the mapping between both coordinate systems by the integral relation

$$x^\rho \to \bar{x}^\rho = \int_0^{x^\rho} \mathrm{d}x^{\rho\prime}\ s_\rho(x^{\rho\prime})\,. \tag{7.35}$$

Both variants are equivalent and connected via the relation

$$\frac{\partial \bar{x}^\rho}{\partial x^\rho} \overset{\text{Eq. (7.34)}}{=} \frac{1}{f_\rho(x^\rho)} = s_\rho(x^\rho)\,. \tag{7.36}$$

For the sake of improved readability, we introduce the abbreviations $f_x = f_1(x^1)$, $f_y = f_2(x^2)$, $s_x = s_1(x^1)$, and $s_y = s_2(x^2)$.

7.2.2. Effective Permittivity and Permeability

Using Eqs. (7.33), the transformation laws for the permittivity, Eqs. (7.3), read

$$\varepsilon^{11} = \sqrt{g}\,\frac{\partial x^1}{\partial \bar{x}^1}\,\frac{\partial x^1}{\partial \bar{x}^1}\,\bar{\varepsilon}^{11} = \sqrt{g}\;f_x\;f_x\;\bar{\varepsilon}^{11}\,, \tag{7.37a}$$

$$\varepsilon^{12} = \sqrt{g}\,\frac{\partial x^1}{\partial \bar{x}^1}\,\frac{\partial x^2}{\partial \bar{x}^2}\,\bar{\varepsilon}^{12} = \sqrt{g}\;f_x\;f_y\;\bar{\varepsilon}^{12}\,, \tag{7.37b}$$

$$\varepsilon^{13} = \sqrt{g}\,\frac{\partial x^1}{\partial \bar{x}^1}\,\frac{\partial x^3}{\partial \bar{x}^3}\,\bar{\varepsilon}^{13} = \sqrt{g}\;f_x\;1\;\bar{\varepsilon}^{13}\,, \tag{7.37c}$$

$$\varepsilon^{21} = \sqrt{g}\,\frac{\partial x^2}{\partial \bar{x}^2}\,\frac{\partial x^1}{\partial \bar{x}^1}\,\bar{\varepsilon}^{21} = \sqrt{g}\;f_y\;f_x\;\bar{\varepsilon}^{21}\,, \tag{7.37d}$$

$$\varepsilon^{22} = \sqrt{g}\,\frac{\partial x^2}{\partial \bar{x}^2}\,\frac{\partial x^2}{\partial \bar{x}^2}\,\bar{\varepsilon}^{22} = \sqrt{g}\;f_y\;f_y\;\bar{\varepsilon}^{22}\,, \tag{7.37e}$$

$$\varepsilon^{23} = \sqrt{g}\,\frac{\partial x^2}{\partial \bar{x}^2}\,\frac{\partial x^3}{\partial \bar{x}^3}\,\bar{\varepsilon}^{23} = \sqrt{g}\;f_x\;1\;\bar{\varepsilon}^{23}\,, \tag{7.37f}$$

$$\varepsilon^{31} = \sqrt{g}\,\frac{\partial x^3}{\partial \bar{x}^3}\,\frac{\partial x^1}{\partial \bar{x}^1}\,\bar{\varepsilon}^{31} = \sqrt{g}\;1\;f_x\;\bar{\varepsilon}^{31}\,, \tag{7.37g}$$

$$\varepsilon^{32} = \sqrt{g}\,\frac{\partial x^3}{\partial \bar{x}^3}\,\frac{\partial x^2}{\partial \bar{x}^2}\,\bar{\varepsilon}^{32} = \sqrt{g}\;1\;f_y\;\bar{\varepsilon}^{32}\,, \tag{7.37h}$$

$$\varepsilon^{33} = \sqrt{g}\,\frac{\partial x^3}{\partial \bar{x}^3}\,\frac{\partial x^3}{\partial \bar{x}^3}\,\bar{\varepsilon}^{33} = \sqrt{g}\;1\;1\;\bar{\varepsilon}^{33}\,. \tag{7.37i}$$

All other derivatives — the mixed terms — in the sum of Eqs. (7.3) vanish. Note that the permittivity $\varepsilon^{\rho\sigma} = \varepsilon^{\rho\sigma}(x^1, x^2, x^3)$ is now a function of the transformed space, whereas $\bar{\varepsilon}^{\rho\sigma}$ are functions of the physical (Euclidean) space. The prefactor stemming from the metric's determinant can be calculated from Eq. (7.7):

$$\sqrt{g} = \frac{1}{f_x\;f_y}\,. \tag{7.38}$$

This leads to the effective permittivity tensor

$$\underline{\boldsymbol{\varepsilon}} \;=\; \begin{pmatrix} \frac{f_x}{f_y} & 1 & \frac{1}{f_y} \\ 1 & \frac{f_y}{f_x} & \frac{1}{f_x} \\ \frac{1}{f_y} & \frac{1}{f_x} & \frac{1}{f_x\,f_y} \end{pmatrix} * \underline{\bar{\boldsymbol{\varepsilon}}} \;=\; \begin{pmatrix} \frac{s_y}{s_x} & 1 & s_y \\ 1 & \frac{s_x}{s_y} & s_x \\ s_y & s_x & s_x\,s_y \end{pmatrix} * \underline{\bar{\boldsymbol{\varepsilon}}} \;\equiv\; \underline{\boldsymbol{\Pi}} * \underline{\bar{\boldsymbol{\varepsilon}}}\,, \tag{7.39}$$

where the $*$ denotes a component wise multiplication. In the same fashion we can easily determine the respective formula for the permeability. Replacing $\underline{\boldsymbol{\varepsilon}} \to \underline{\boldsymbol{\mu}}$, and for non-magnetic materials $\underline{\bar{\boldsymbol{\mu}}} \to \mathbb{1}$ in Eq. (7.39), we finally arrive at

$$\underline{\boldsymbol{\mu}} \;=\; \begin{pmatrix} \frac{f_x}{f_y} & 0 & 0 \\ 0 & \frac{f_y}{f_x} & 0 \\ 0 & 0 & \frac{1}{f_x\,f_y} \end{pmatrix} \;=\; \begin{pmatrix} \frac{s_y}{s_x} & 0 & 0 \\ 0 & \frac{s_x}{s_y} & 0 \\ 0 & 0 & s_x\,s_y \end{pmatrix} \;\equiv\; \underline{\boldsymbol{\Pi}} * \underline{\mathbf{1}}\,. \tag{7.40}$$

The missing piece is the definition of appropriate mapping functions F_ρ or the related functions f_ρ and s_ρ. Therefore, we implement two different PML types in the following paragraphs.

7.2.3. Lalanne Formulation

The credit for the introduction of stretched-coordinate PMLs into the FMM realm can be clearly attributed to Lalanne *et al.* [99]. They suggested a mapping function which squeezes the infinite complex space onto a finite unit cell such that the artificial periodicity can be maintained. The mapping bases on a tangent like function since these functions are known to diverge for arguments approaching $\pm\pi/2$.

The mapping function of Lalanne *et al.*is given as [99]

$$F_\rho(x^\rho) = \begin{cases} x^\rho, & \text{for } |x^\rho| \leq \frac{e}{2}, \\ \frac{x^\rho}{|x^\rho|}\left(\frac{e}{2} + \frac{q}{\pi(1-\gamma)}\left(\tan(\phi) - \frac{\gamma}{\sqrt{1-\gamma}}\tan^{-1}\left(\sqrt{1-\gamma}\tan(\phi)\right)\right)\right), & \text{for } \frac{e}{2} < |x^\rho| < \frac{d}{2}, \end{cases} \tag{7.41a}$$

and the corresponding inverse derivative as

$$f_\rho(x^\rho) = \begin{cases} 1, & \text{for } |x^\rho| \leq \frac{e}{2}, \\ \left(1 - \gamma\sin^2(\phi)\right)\cos^2(\phi), & \text{for } \frac{e}{2} < |x^\rho| < \frac{d}{2}, \end{cases} \tag{7.41b}$$

with $\phi = \frac{\pi}{q}\left(|x^\rho| - e/2\right)$, $q = (d-e)$, and $\gamma \in \mathbb{C}$. The functions of Eqs. (7.41) are visualized in Fig. 7.14 for $\gamma = (0.5 + 0.5\mathrm{i})$, and are split into real and imaginary part. Note that we usually pick d slightly larger than the lattice constant a, which is a necessity of our implementation to avoid divisions by zero in Eq. (7.39) and Eq. (7.40).

This choice of the mapping function has the advantage that the Fourier coefficients of f_ρ can be

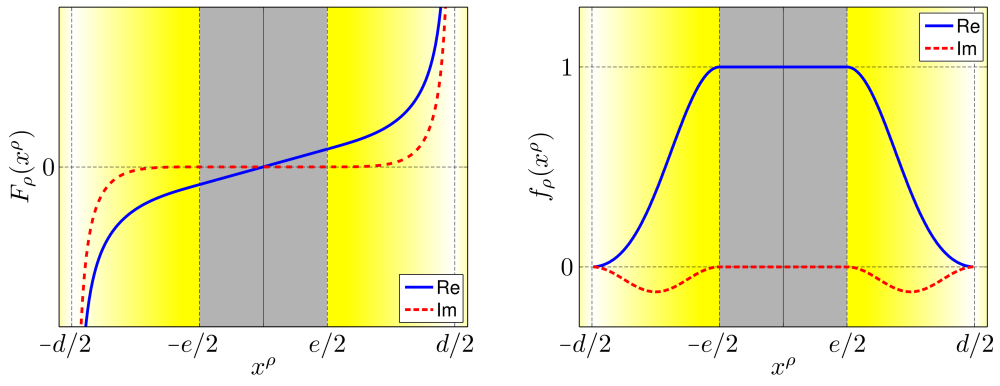


Figure 7.14.: *Complex stretched-coordinate transformation mapping function F_ρ and its inverse derivative f_ρ as proposed for PML applications by Lalanne et al. [99]. The parameters chosen for the picture are $e = 0.33$ and $\gamma = (0.5 + 0.5\mathrm{i})$.*

determined analytically. They read

$$\tilde{f}_{\rho,m} = \delta_{0m} - \frac{q}{2d}(-1)^m \left[\left(1+\frac{\pi}{4}\right)\mathrm{sinc}\left(\frac{mq}{d}\right) + \frac{1}{2}\mathrm{sinc}\left(\frac{mq}{d}-1\right) + \frac{1}{2}\mathrm{sinc}\left(\frac{mq}{d}+1\right) + \frac{\gamma}{8}\mathrm{sinc}\left(\frac{mq}{d}-2\right) - \frac{\gamma}{8}\mathrm{sinc}\left(\frac{mq}{d}+2\right)\right], \tag{7.42}$$

with $\mathrm{sinc}(x) = \sin(\pi x)/(\pi x)$. We notice that the Fourier coefficients of a sinc decay only slowly with $\frac{1}{m}$. Furthermore, the peaks of the last two sinc terms are positioned at $(\frac{mq}{d} \pm 2) = 0$. This means, if the PML region takes about as much as $\frac{q}{d} = 4/10$ of the unit cell (which is usually fairly large), the coefficients' values up to $m = \pm 5$ are still in the order of one. Consequently, the corresponding Fourier series is expected to converge rather slowly. This might not be a problem for one-dimensional PML calculations (two-dimensional problems, with artificial periodicity in one transverse direction) as used in Ref. [99], where truncation orders of $m \approx \pm 500$ are easily within reach. However, in three-dimensional problems (two PML dimensions) the maintainable truncation order for realistic applications is about $m \approx \pm 15 \ldots 20$ depending on the truncation scheme.[16] Considering the convergence behavior of the sinc's Fourier coefficients, it is somewhat surprising that the Fourier coefficients of Eq. (7.42) actually decrease with m^{-3} as can be seen in Fig. 7.16 in the left panel. Still, there remains a relatively large plateau for small m. We attribute this plateau mainly to the constant part of f_ρ in the interval $[-e/2, e/2]$ which is not easily represented in a Fourier series.[17] Hence, for 3D applications with small truncation orders per dimension, we expect a rather slow convergence. What happens when we reconstruct the inverse derivative f_ρ from the Fourier coefficients of Eq. (7.42) with a truncation order of $m = \pm 20$ is depicted in Fig. 7.15. The consequence of this truncated Fourier series are spurious residual oscillations in the order of 10^{-3}. The smaller the constant parts of the physical domain in relation to the unit cell size, the smaller the

[16] Please note that the eigenproblem diagonalization scales with the number of coefficients to the third, which means roughly with the truncation order to the sixth.

[17] This is not a problem for any polynomial expansion basis though.

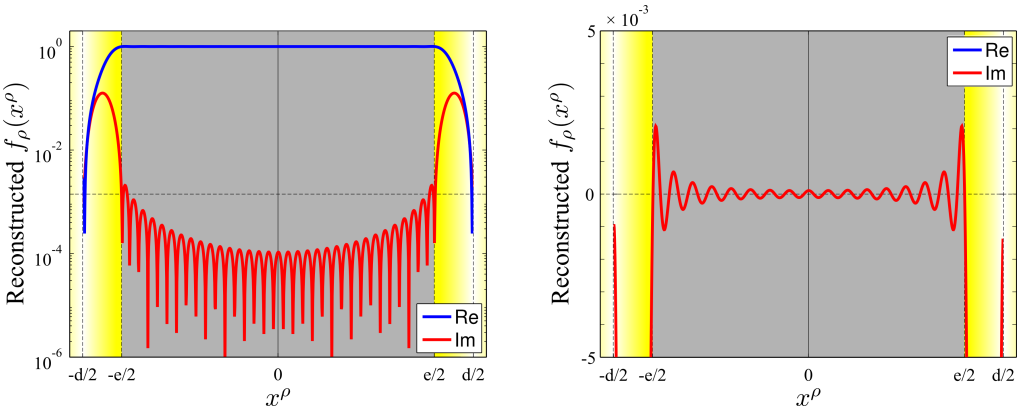


Figure 7.15.: *Fourier reconstructed inverse derivative $f_\rho(x^\rho)$ of the Lalanne mapping function for a truncation order of $m = \pm 20$ (one-dimensional, $M = 41$) and $e = 0.8$. The left panel shows the magnitude of the reconstructed function on a logarithmic scale, and the right panel focuses on the imaginary part on a linear scale. Note the oscillations around zero within the physical domain (grey shaded).*

plateau, because the sinc functions move towards the origin. For instance, for $e = 0.33$, when the PML domain is larger than the physical domain, the real as well as the imaginary parts move below the red curve. Furthermore, the magnitude of the residual spurious oscillations drops for about one order of magnitude in this case. Indeed, we will see effects related to this convergence behavior and consequences for the eigenmodes in the detailed analysis of the PML performance in Chap. 8.

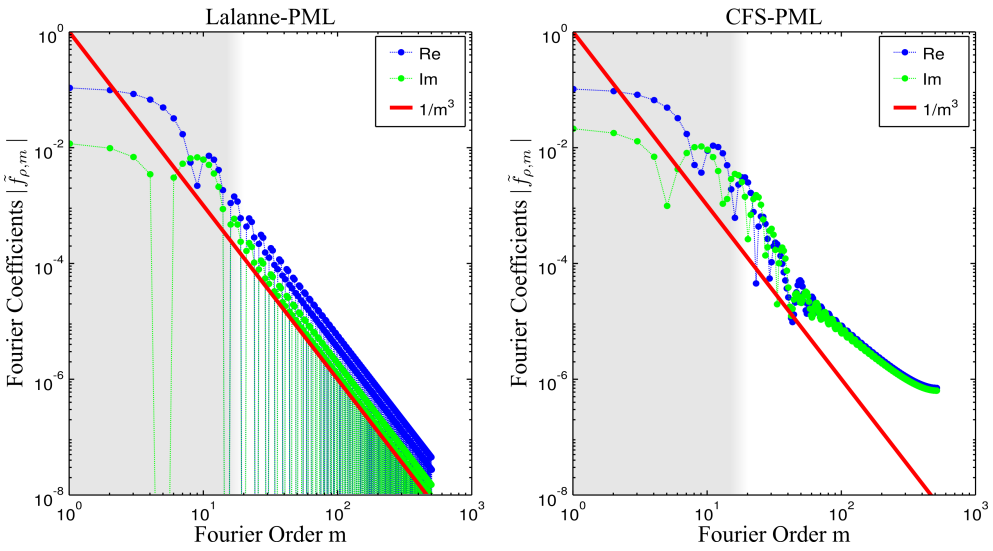


Figure 7.16.: *Convergence of Fourier coefficients $\tilde{f}_{\rho,m}$ of the inverse derivative f_ρ for Lalanne formulation (left), and third-order polynomial CFS formulation (right). The coefficients of the Lalanne formulation are those of Eq. (7.42), whereas the coefficients of the CFS formulations were determined by an FFT with $N_{\text{fft}} = 1024$ sampling points (hence, the small aliasing effect at large m). The grey shaded area roughly highlights the coefficients that can be taken into account in 3D simulations. The used transformation parameters are $e = 0.8$,$\gamma = (0.5+0.5\text{i})$, $\omega = 1$, $\kappa_{\max} = 10$, $m_\kappa = 3$, $\sigma_{\max} = 10$, $m_\sigma = 3$, $a_{\max} = 0$, and $m_a = 1$.*

7.2.4. Complex Frequency Shifted Formulation with Polynomial Grading

A PML type commonly used in real–space methods is the complex frequency shifted PML (CFS-PML) which brings sufficient degrees of freedom to construct a causal PML medium [105, 112]. We implemented this specific type of PML because with its general form it can easily mimic other kinds of less sophisticated PMLs like, for example, uniaxial perfectly matched layers (UPML) by proper choice of the parameters as well. Originating from real–space methods where polynomials are the most widespread type of basis functions, the stretched mapping within the edge domain is achieved by a polynomial grading of a freely selectable degree.

In a material independent formulation with dimensionless conductivities σ,[18] and an exponential time dependence of $e^{-\text{i}\omega t}$ (responsible for the red minus sign in Eq. (7.43a)), the slope s_ρ of the

[18] We pick: $\sigma = (\sigma_{\text{SI}} a)/(c\varepsilon\varepsilon_0)$.

mapping function F_ρ is given by

$$s_\rho(x^\rho) = \begin{cases} \kappa_\rho(x^\rho) - \frac{\sigma_\rho(x^\rho)}{a_\rho(x^\rho)+\mathrm{i}\omega} = \left(\kappa_\rho(x^\rho) - \frac{a_\rho\sigma_\rho}{a_\rho^2+\omega^2}\right) + \mathrm{i}\,\frac{\omega\sigma_\rho}{a_\rho^2+\omega^2}\,, & \text{for } |x^\rho| \in \left[\frac{e}{2}, \frac{d}{2}\right], \\ 1, & \text{else,} \end{cases} \tag{7.43a}$$

with polynomially graded parameters

$$\kappa_\rho(x^\rho) = 1.0 + (\kappa_{\max} - 1.0) \cdot \left(\frac{(2|x^\rho| - e)}{q}\right)^{m_\kappa}, \tag{7.43b}$$

$$\sigma_\rho(x^\rho) = \sigma_{\max} \cdot \left(\frac{(2|x^\rho| - e)}{q}\right)^{m_\sigma}, \tag{7.43c}$$

$$a_\rho(x^\rho) = a_{\max} \cdot \left(\frac{(d - 2|x^\rho)|}{q}\right)^{m_a}. \tag{7.43d}$$

According to Ref. [105], this transformation maintains causality provided we choose $\kappa_\rho \geq 1$, $\sigma_\rho \geq 0$, and $a_\rho \geq 0$ and real.

In order to get the primitive $F_\rho(x^\rho)$, we must integrate Eq. (7.43a). For a vanishing frequency shift $a_\rho = 0$, Eq. (7.43a) reduces to $s_\rho = \kappa_\rho + \mathrm{i}\sigma_\rho/\omega$, and the mapping function becomes

$$F_\rho(x^\rho) = \begin{cases} x^\rho, & \text{for } |x^\rho| \leq \frac{e}{2}, \\ \frac{x^\rho}{|x^\rho|}\left(|x^\rho| + \frac{\kappa_{\max}-1}{m_\kappa+1}\left(\frac{2}{q}\right)^{m_\kappa}\tilde{x}^{m_\kappa+1} + \mathrm{i}\frac{\sigma_{\max}}{\omega(m_\sigma+1)}\left(\frac{2}{q}\right)^{m_\sigma}\tilde{x}^{m_\sigma+1}\right), & \text{for } |x^\rho| \in \left[\frac{e}{2}, \frac{d}{2}\right], \end{cases} \tag{7.44}$$

with the abbreviation $\tilde{x} = \left(|x^\rho| - \frac{e}{2}\right)$. The mapping and the corresponding inverse derivative for a third order polynomial grading are plotted in Fig. 7.17. The difference in comparison to the Lalanne mapping, see Fig. 7.14, is evident: Where the tangent-like function in Eq. (7.41a) maps the *infinite*

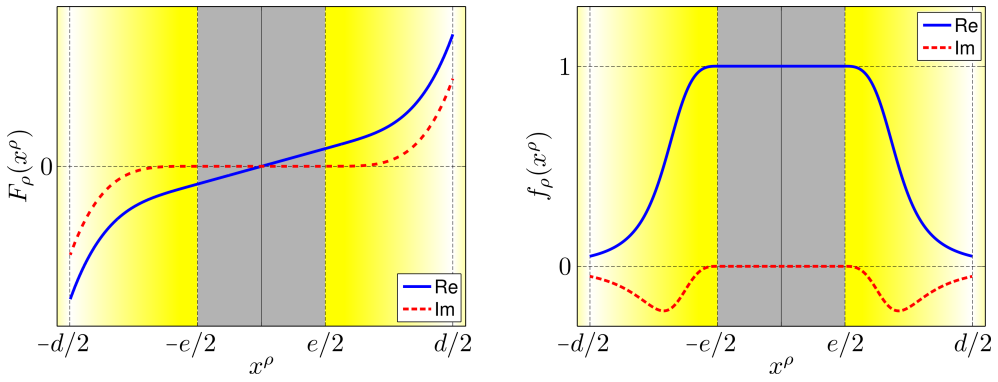


Figure 7.17.: *Complex stretched-coordinate transformation mapping function F_ρ and its inverse derivative f_ρ as proposed for PML applications in CFS formulation. The parameters chosen for the picture are $\omega = 1$, $e = 0.33$, $\kappa_{\max} = 10$, $m_\kappa = 3$, $\sigma_{\max} = 10$, $m_\sigma = 3$, $a_{\max} = 0$, and $m_a = 1$.*

physical space onto the finite PML region, the CFS-PML only squeezes a *finite* domain of the physical space onto the edge layer. How large this region in physical space is, is determined by the slope of the real part $\kappa_\rho(x^\rho)$ integrated over the PML thickness. It can be adjusted by the maximal slope $\kappa_{\max}$. Similarly, the attenuation strength is controlled by the maximal slope of the imaginary part — the artificial conductivity $\sigma_{\max}$.

It is not much surprising that the Fourier coefficients of the polynomial CFS mapping do not decrease much faster than those of the Lalanne mapping, as can be seen in the right panel of Fig. 7.16. After all, we attributed the plateau for small orders to the constant part in the physical domain which is the same in both functions. It seems that the polynomial grading is slightly worse to expand into plane waves and the convergence order is considerably smaller than -3. Hence, we expect in general no better performance from polynomial CFS-PMLs. This expectation could be confirmed within the numerical experiments that were evaluated in the course of this work. In all experiments their performance was similar to infinite Lalanne PMLs. However, it might be worthwhile to examine the performance more rigorously in future projects.

7.3. Combination of Coordinate Transformations

One of the great advantages of the covariant formulation of Maxwell's equations is that we can easily combine several coordinate transformations. We already mentioned that adaptive coordinates and adaptive spatial resolution are most commonly done together. However, the combination with perfectly matched layers, to the best of our knowledge, has not been reported in literature before. In this section we demonstrate the formalism of how this can be achieved.

The combination of ASR and AC affects only the sampling points at which the effective permittivity is evaluated. If

$$\chi_{kl} = (x_k^1, x_l^2) \tag{7.45}$$

denotes the equidistant sampling points of a cartesian grid with $x_k^1 = \Delta_1 k$, $x_l^2 = \Delta_2 l$, $k, l = 0, \ldots, N_{\text{fft}} - 1$, the ASR maps them to the new points

$$\hat{\chi}_{kl}(\chi_{kl}) = (\hat{x}_k^1(x_k^1), \hat{x}_l^2(x_l^2))\,, \tag{7.46}$$

where the new coordinates

$$\hat{x}^1(x^1) = \mathrm{GV}(x^1, \{(p_l, \hat{p}_l)\}^1, G)\,, \tag{7.47a}$$

$$\hat{x}^2(x^2) = \mathrm{GV}(x^2, \{(p_l, \hat{p}_l)\}^2, G)\,. \tag{7.47b}$$

are given by the generalized Vallius transformation GV introduced in Sec. 7.1.1 (cf. Ref. [100]). Here, $\{(p_l, \hat{p}_l)\}^\rho$ denotes the set of inflection points for the respective direction. Note that the newly introduced coordinates $\hat{x}_k^1$ and $\hat{x}_l^2$ simply describe the new, non-Cartesian ASR mesh in transformed space ($\mathcal{O}x^1x^2x^3$). Hence, the discretized permittivity in transformed space is given by

$$\varepsilon^{\rho\sigma}(\chi_{kl}) \overset{\text{Eqs. (7.3)}}{=} \sqrt{g(\hat{\chi}_{kl})} \left(\frac{\partial x^\rho}{\partial \bar{x}^\tau} \frac{\partial x^\sigma}{\partial \bar{x}^\kappa} \right) \Bigg|_{\hat{\chi}_{kl}} \bar{\varepsilon}^{\tau\kappa}(\bar{\chi}_{kl})\,, \tag{7.48}$$

with

$$\bar{\chi}_{kl}(\hat{\chi}_{kl}) = (\bar{x}^1(\hat{x}_k^1), \bar{x}^2(\hat{x}_l^2))\,. \tag{7.49}$$

Here, $\bar{\varepsilon}^{\tau\kappa}$ is the permittivity within the unit cell of a (infinitely) periodic system in physical space. Equation (7.48) can subsequently be transformed into Fourier space.

Due to the lack of mutual dependence in the PML transformation function, the additional inclusion of open boundaries is not much different. We consider a structure within infinite free space — the original physical problem. The essential additional step is to squeeze the infinite space onto a finite unit cell using PMLs. If the original aperiodic permittivity is described by $\bar{\varepsilon}^{\tau\kappa}$,[19] the discretized permittivity in transformed space is given by

$$\varepsilon^{\rho\sigma}(\chi_{kl}) = \sqrt{g(\hat{\chi}_{kl})}\,\left(\frac{\partial x^{\rho}}{\partial \bar{x}^{\tau}}\frac{\partial x^{\sigma}}{\partial \bar{x}^{\kappa}}\right)\Bigg|_{\hat{\chi}_{kl}} \underbrace{\left(\Pi^{\tau\kappa}(\hat{\chi}_{kl})\;\bar{\varepsilon}^{\tau\kappa}(\bar{\chi}_{kl})\right)}_{\text{PML transformed permittivity}}, \tag{7.50}$$

where $\Pi^{\tau\kappa}$ are the components of matrix $\underline{\mathbf{\Pi}}$ defined in Eq. (7.39). The rightmost term in brackets is just the permittivity in the squeezed space of the finite unit cell from Eq. (7.39), which replaces the periodic permittivity from Eq. (7.48). Actually, the permittivity $\bar{\varepsilon}^{\tau\kappa}(\bar{\chi}_{kl})$ is here an analytic continuation of the real-space permittivity function to complex space. It is essentially equivalent to $\bar{\varepsilon}^{\tau\kappa}(\mathrm{Re}(\bar{\chi}_{kl}))$. The transformation of the permeability is carried out analogously.

7.4. Transformation into Fourier Space

There are two slightly different procedures available for the transformation into Fourier space connected to two different general strategies. These two strategies and their differences are briefly discussed in the following. We reduce the considerations to the permittivity. It is understood that the permeability is treated analogously.

7.4.1. Structure-Transform Real-Space Strategy

The derivations in the first section of this chapter are based on the covariant formulation of Maxwell's equations. This means, instead of an incorporation of the coordinate transformations into Maxwell's equations itself, we handle the coordinate transformations by altering the calculated structure. We demonstrated above that, in this way, we can treat AC, ASR, or PML transformations with the same formalism. The presented formalism takes care of coupling of various tensor components due to non-separable multi-dimensional transformations. Furthermore, it was shown that it is rather easy to combine different transformations without altering the form of the eigenvalue problem. The effects of ASR, AC, and PML could be converted into two final anisotropic effective structure functions given in continuous real-space in one step. Hence, we call this procedure the *structure-transform real-space strategy*.

The discretization and transformation into Fourier space happens only in a last single step. As already described in Eqs. (6.10), the combined effective material parameters, i.e., Eq. (7.48) or Eq. (7.50), are Fourier transformed using Li's Fourier transformation operators

$$\hat{\underline{\boldsymbol{\varepsilon}}} = \hat{\mathrm{l}}_3^{-}\,\frac{\hat{\mathrm{L}}_2\hat{\mathrm{L}}_1 + \hat{\mathrm{L}}_1\hat{\mathrm{L}}_2}{2}\,\underline{\boldsymbol{\varepsilon}}(\chi_{kl})\,. \tag{7.51}$$

[19] Note the fundamental difference to the physical system described before!

The Fourier representations $\hat{\underline{\varepsilon}}$ of the effective permittivity in transformed space can readily be used within the discretized system matrix, Eq. (6.12). The basis functions of the FMM are then plane waves in the transformed space. The obtained fields reside in the transformed space as well.

7.4.2. Equation-Transform k-Space Strategy

The second — analytically equivalent — strategy we call *equation-transform k-space strategy*. This strategy is, for example, used by Lalanne *et al.* [99] in their original proposal of SC-PMLs for the FMM. If we do not stick to the covariant formulation of Maxwell's equations, but derive the FMM formalism directly from the dimensionless curl equations, Eqs. (2.13), every coordinate transformation alters their form and adds some new terms. Coordinate transformations enter Maxwell's curl equations via the derivatives

$$\frac{d}{d\bar{x}^\rho} \quad \longrightarrow \quad \frac{dx^\sigma}{d\bar{x}^\rho}\frac{d}{dx^\sigma}\,. \tag{7.52}$$

The structure functions $\underline{\varepsilon}$ and $\underline{\boldsymbol{\mu}}$, however, remain unchanged. For PMLs we have separately implemented this strategy. Because of the independent transformations in both transverse dimensions, the derivatives in physical space

$$\frac{d}{d\bar{x}^\rho} \quad \xrightarrow{\text{SC-PML}} \quad f_\rho(x^\rho)\frac{d}{dx^\rho} \tag{7.53}$$

can be substituted by the derivatives in squeezed space times the inverse derivative f_ρ, for $\rho = 1, 2$. If we take as example the first component of Eq. (2.13a)

$$\bar{\partial}_2\bar{E}_3 - \bar{\partial}_3\bar{E}_2 = \mathrm{i}\omega^2\left(\bar{\mu}_{11}\bar{H}_1 + \bar{\mu}_{12}\bar{H}_2 + \bar{\mu}_{13}\bar{H}_3\right)\,, \tag{7.54a}$$

(in non-covariant form), the PML transformed equation within the finite unit cell reads

$$f_2\partial_2 E_3 - \partial_3 E_2 = \mathrm{i}\omega^2\Big(\mu_{11}H_1 + \mu_{12}H_2 + \mu_{13}H_3\Big)\,, \tag{7.54b}$$

with $\mu_{\rho\sigma}(x^1, x^2) = \bar{\mu}_{\rho\sigma}(F_1(x^1), F_2(x^2))$. The remaining components are transformed similarly and combined to an eigenvalue equation. The main difference to the structure-transform real-space strategy is that material parameters and terms from the derivatives are now Fourier transformed separately obeying Li's product rules, Eq. (3.26) and Eq. (3.29), in order to be able to easily execute the remaining derivatives. The products of real-space functions become convolutions in Fourier space, described by the products of Toeplitz matrices and Fourier vectors. This means that Eq. (7.54b) in Fourier space is given by

$$[\![f_2]\!]\underline{\boldsymbol{\beta}}\,\tilde{\mathbf{E}}_3 - \gamma\tilde{\mathbf{E}}_2 = \omega^2\Big([\![\mu_{11}]\!]\tilde{\mathbf{H}}_1 + [\![\mu_{12}]\!]\tilde{\mathbf{H}}_2 + [\![\mu_{13}]\!]\tilde{\mathbf{H}}_3\Big)\,. \tag{7.55}$$

Here, $\underline{\boldsymbol{\beta}}$ describes a diagonal matrix with entries $\underline{\boldsymbol{\beta}}_{mn} = \beta_n$ (similarly $\underline{\boldsymbol{\alpha}}_{mn} = \alpha_m$, cf. Eqs. (3.47)). The corresponding small eigenproblem system matrices $\underline{\mathbf{F}}$ and $\underline{\mathbf{G}}$ for non-magnetic, isotropic systems can be found in App. B.4. More details on the eigenproblem in the equation-transform k-space strategy can be found in Ref. [84]. In summary, this strategy transforms Maxwell's equations directly and adds the coordinate transformation and PML effect as convolutions in k-space.

7.4.3. Differences

While both strategies are equivalent on an analytical level, there are subtle differences in the numerical implementation. We would like to name two differences which stand out.

The first difference is the point at which the discretization and transformation into Fourier space is carried out. The structure-transform real-space strategy constructs effective material functions in real-space as an aggregate of all involved functions on an exact level first. The discretization (and, therefore, discretization errors) is only introduced once in the end. Furthermore, only the discretized effective permittivity and permeability are Fourier transformed and convolved with the fields. Hence, the infinite Fourier series is truncated once. However, the Li operators do some non-trivial mixing and convolution of the components to enhance the convergence behavior. In contrast, the equation-transform k-space strategy carries out the real-space discretization and Fourier transformation for every single involved transformation function. All truncated series are then convolved one after the other taking Li's rules into account.

The second difference concerns the nature of the Fourier transformed functions, especially if we apply PML transformations.The latter strategy only transforms functions f_ρ whose real parts are bounded on the interval $\mathrm{Re}(f_\rho) \in [0, 1]$, and whose imaginary parts are roughly bounded on the interval $\mathrm{Im}(f_\rho) \in [-0.5, 0]$. When we have a close look at Eq. (7.39) and Eq. (7.40) we notice that the former strategy, however, Fourier expands terms involving $1/f_\rho = s_\rho$. These functions can have quite large values and steep shoulders. For the Lalanne SC-PML these terms even tend to diverge as f_ρ approaches zero at the outer unit cell boundaries. Their truncated Fourier series are expected to converge considerably slower than those of f_ρ in the equation-transform k-space strategy. These inverse terms are introduced by the square root of the metric determinante $\sqrt{g}$ (cf. Eq. (7.38)) which originates from the curl operator (cf. Eq. (2.80)). This means, during the covariant formulation of Maxwell's equations we actually created those terms by multiplication with $\sqrt{g}$ in order to absorb this spatially dependent term into the effective permittivity.

7.5. Back-Transformation into Cartesian Space

The field components of solutions calculated with the coordinate transformation schemes above can be obtained from Eq. (6.30) and Eqs. (6.31) as usual. However, they are given with respect to the plane wave basis in transformed space. In order to obtain the Cartesian fields in physical space, the components have to be transformed back. The transformation process in general involves two steps. We present the procedure for the electric fields — the magnetic fields follow in the same way.

The first step is the back transformation of the curvilinear field components into the Cartesian ones. To this end, we apply Eq. (2.90) in Fourier representation, which reads

$$\tilde{\bar{\mathbf{E}}}^{(l)}_{\rho'}(\bar{x}^3(x^3)) = \left[\!\left[\bar{\Lambda}^{\rho}{}_{\rho'} \right]\!\right] \tilde{\mathbf{E}}^{(l)}_{\rho}(x^3)\,. \tag{7.56}$$

Here, we reintroduced the layer label (l) for completeness. Please also note that the Einstein summation convention is implicitly assumed again. Since these Fourier coefficients were created with respect to the transformed coordinates x^1 and x^2, the field values obtained from an inverse FFT are the Cartesian field components $\bar{E}_{\rho'}(\bar{\chi}_{kl}, \bar{x}^3)$ on the adapted mesh. In many cases this is a rather

convenient representation because, plotted on this mesh in Cartesian space, the pictures still have the exact surface representation and the locally enhanced resolution from the ASR.

If the field values are required in the plane wave basis of the Cartesian space or have to be plotted on the Cartesian mesh an additional step is necessary. Then, we have to transform the plane wave basis of transformed space back into the plane wave basis of Cartesian space which requires a transformation matrix $\underline{\boldsymbol{L}}$ with entries [65]

$$L_{mn} = \int_0^1 \int_0^1 \mathrm{d}\bar{x}^1 \mathrm{d}\bar{x}^2 \; e^{-\mathrm{i}(\bar{\alpha}_{m_1}\bar{x}^1 + \bar{\beta}_{m_2}\bar{x}^2)} \; e^{-\mathrm{i}(\alpha_{n_1}x^1 + \beta_{n_2}x^2)} \,. \tag{7.57}$$

The indices m and n denote multi-indices as defined in Chap. 3. Due to the integration, the preparation of this matrix is rather cumbersome. The fields in the Cartesian plane wave basis are then given by

$$\tilde{\bar{\mathbf{E}}}^{(l)}_{\rho'}(\bar{x}^3(x^3)) = \underline{\boldsymbol{L}} \left[\!\!\left[\bar{\Lambda}^{\rho}{}_{\rho'} \right]\!\!\right] \tilde{\mathbf{E}}^{(l)}_{\rho}(x^3) \,, \tag{7.58}$$

from which the real-space fields on the Cartesian grid $\bar{E}_{\rho'}(\bar{x}^1_k, \bar{x}^2_l, \bar{x}^3)$ can be obtained by means of the field reconstruction methods presented in Sec. 6.7.

An exception to this procedure are PML transformations, which are never reverted. Instead, the solutions are only evaluated within the central physical domain where the identity transformation was applied. Hence, the fields in this domain remain physical even in squeezed space.

Chapter 8.

8 Method Validation

In the previous two chapters we have introduced the Fourier modal method and its extension to adaptive coordinates, adaptive spatial resolution and open boundary conditions on a formal level. The topic of this chapter is a validation and bench marking of these techniques by comparison to analytically obtained reference solutions. Since the introduced extensions are by and large an unworked field of research, our particular interest is the examination and determination of their limits and general rules for a safe use.

As analytical reference solutions for a method based on eigenmode expansion within layered systems, the guided eigenmodes of a cylindrical step index fiber seem worthwhile. The propagation constant or the related effective refractive index are directly accessible and physically significant quantities to compare with. These solutions have already been derived in Sec. 4.1.1. The circular cross sections of the fibers are naturally a challenge to ordinary FMM due to the in-layer staircasing of the Cartesian mesh. Hence, they provide an excellent test for the performance improvement capabilities of AC meshes.

In this chapter, we consider two related physical systems. The first is an artificial periodic arrangement of step-index fibers similar as depicted in Fig. 7.12. If the lattice constant is chosen sufficiently large, the evanescent tails of the guided modes in the cladding region will barely disturb the solutions in neighboring unit cells. These periodic systems are examined in Sec. 8.2 mainly to compare the performance of the established non-differentiable, smoothed, and differentiable meshes.

Thereafter, we isolate the unit cells with additional PMLs and examine those systems. In particular, we describe an optimization scheme for an (non-differentiable) adapted mesh of the step-index fiber. The choice of parameters in this part of the chapter and the selected tests already focus on a further use of the results with the liquid crystal based long period grating mode coupler which is one of the presented applications in Chap. 9. In the last part of this chapter, we stress the challenges of eigenmode calculations with strongly deformed and compressed ASR and PML meshes.

Before we proceed, however, we first address an important basic question in the FMM: What is the influence of the structure size to lattice constant ratio on the overall accuracy — and, hence, what is the appropriate lattice constant for an aperiodic structure?

8.1. Choice of an Appropriate Lattice Constant

There is a vast amount of parameters in the upcoming simulations to be optimized in order to get the best accuracy in the results. Not all of them will be discussed here. However, one of the most basic and also most important ones is the size of the unit cell a. Hence, we want to start our analysis with an investigation of how the lattice constant influences the accuracy of the result. Figure 8.1 shows the relative error of the numerically determined effective refractive indices

$$\mathrm{err}_{\mathrm{rel}}(\mathrm{n}_{\mathrm{eff}}) = \left| \frac{\mathrm{Re}\left(\mathrm{n}_{\mathrm{eff}}\right) - n_{\mathrm{eff},ref}}{n_{\mathrm{eff},ref}} \right| \tag{8.1}$$

for all guided modes of a circular step-index fiber (no jacket, no cladding/air interface, setup details see Sec. 8.3.1) plotted over the lattice constant. Here, "Re" denotes the real part and the subscript "ref" labels the reference solutions obtained as described in Sec. 4.1.1. The total number of plane waves used in the FMM's field expansion is fixed to $M = 997$, and the simulation is performed for $\lambda = 1.25\,\mu\mathrm{m}$.

Figure 8.1(a) depicts the error for the case when the PMLs are switched off, which means that we simulate a periodic lattice of cores in a homogeneous silica background material. Figure 8.1(b) shows the error when the PMLs are switched on (Lalanne's PML, $d = 1.001$, $e = 0.8$, $\gamma = 0.5 + 0.5\mathrm{i}$). When a is varied, we obtain a non-steady contribution to the error from the changing discretization of the core. In order to get rid of this effect, we calculate the core's *effective mean radius* from the discretized permittivity distribution for each value of a separately, and use it to calculate a corresponding reference solution $n_{\mathrm{eff,ref}}(a)$ as described above. The effective mean radii of the core vary

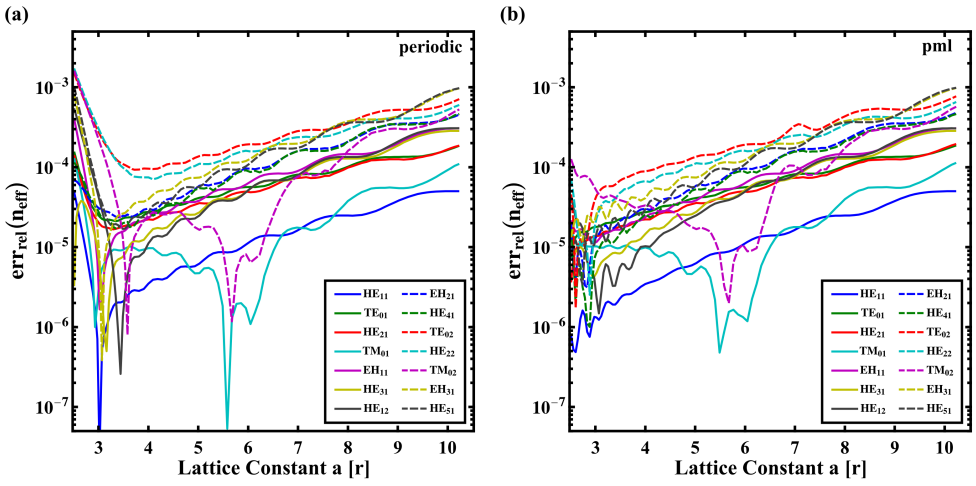


Figure 8.1.: *The relative error of* $\mathrm{n_{eff}}$ *depends on the lattice constant given in units of the radius* r *for an (a) unisolated periodic, and (b) PML isolated unit cell. The number of used plane waves is* $M = 997$ *and the simulated wavelength is* $\lambda = 1.25\,\mu\mathrm{m}$*. The optimal choice for the size of the unit cell in case (a) is around* $a = 3.7r$ *and in case (b) around* $a = 3r$*. Notice the absence of the increasing error for small lattice constants in the PML case, where in the periodic case the error is dominated by lattice effects.*

between $r_{\mathrm{eff,min}} = \min(r_{\mathrm{eff}}(a)) = 2.1498799\,\mu\mathrm{m}$ and $r_{\mathrm{eff,max}} = \max(r_{\mathrm{eff}}(a)) = 2.1506677\,\mu\mathrm{m}$. Note that we do this "correction" in this special case only to carve out the pure influence of the lattice constant. As long as the lattice constant stays fixed in simulations later on, the discretization error is a contribution to the total error which solely depends on the chosen (adapted) mesh and the number of sampling points (cf. Sec. 8.2). Hence, the remaining errors plotted in Fig. 8.1 *exclude* most influences from an imperfect discretization which are common for the method. Usually, these discretization errors can be minimized by a proper choice of the number of sampling points N_{fft} for the Fast Fourier Transform (FFT). The following simulations are always performed with $N_{\mathrm{fft}} = 1024$ sampling points in each transverse direction.

The results for periodic boundary conditions in Fig. 8.1(a) show that, at least for the guided modes which decay exponentially outside of the core region, the standard FMM approach yields good results if the unit cell size is sufficiently large. Sufficiently large in this regard means that the results get better the smaller the evanescent tails of the fields are at the unit cell boundary (or at least at the neighboring core). Unfortunately, this *positive* convergence effect for growing a is overcompensated by a countering convergence effect from a decreasing structure size to unit cell size ratio. The latter *negative* convergence effect occurs for increasing lattice constants at a constant number of plane waves as well. It can be explained by the fact that for increasing unit cell size the reciprocal lattice vectors become shorter. Thus, even though the k-space resolution gets finer, the area in k-space covered by a certain finite number of reciprocal lattice vectors decreases as well. Both oppositional effects entail an optimal unit cell size for a given wavelength and number of plane waves, here at around $a = 3.7r$.

As can be concluded from comparing Fig. 8.1(a) and Fig. 8.1(b), for large lattice constants the PMLs barely influence the guided modes' propagation constants — the limited number of reciprocal lattice vectors used seems to dominate the error as before. In contrast, for small lattice constants the error with PMLs is much smaller. Here, the stretched coordinate transform of the PML with the enhanced attenuation of the evanescent tails contributes to the accuracy. However, the error does not entirely vanish for small lattice constants.

On the one hand, this can be attributed to the still limited k-space resolution and coverage. On the other hand, a great part of the remaining deviation from the reference could also be due to the effect that the absorption in the PML region eventually not only leads to a stronger attenuation of the evanescent fields, but also to additional rapid oscillations with decreasing period lengths when reaching the outer boundary, especially for SC-PMLs where both real and imaginary part of the coordinate reach large values at the boundary. This is an issue we have not discussed in Sec. 7.2. Analytically, the period length[1] approaches zero at those points. This becomes plausible when reconsidering Eq. (7.31). Now, assuming *both* k and x to represent a complex wave vector and a complex streched coordinate in transverse direction, respectively, the one-dimensional plane wave can be decomposed into

$$e^{\mathrm{i}kx} \quad \longrightarrow \quad e^{\mathrm{i}(k_{\mathrm{R}}+\mathrm{i}k_{\mathrm{I}})(x_{\mathrm{R}}+\mathrm{i}x_{\mathrm{I}})} \;=\; e^{\mathrm{i}k_{\mathrm{R}}x_{\mathrm{R}}}\; e^{-k_{\mathrm{R}}x_{\mathrm{I}}}\; e^{-k_{\mathrm{I}}x_{\mathrm{R}}}\; e^{-\mathrm{i}k_{\mathrm{I}}x_{\mathrm{I}}}\,. \tag{8.2}$$

For evanescent tails of guided modes (transverse evanescent, $k_{\mathrm{R}} = 0$), the rapid oscillations stem from the fourth exponential term on the right-hand side with $x_{\mathrm{I}} \to \infty$. Usually, with the available numbers of degrees of freedom, these oscillations cannot be resolved properly. A similar argument holds for radiative solutions (transverse propagating, $k_{\mathrm{I}} = 0$), what we can see from the first exponential term on the right-hand side of Eq. (8.2) with the difference that the real part of the coordinate

[1] Which can be seen as being proportional to f_ρ.

stretching is responsible for the rapid oscillations ($x_\mathrm{R} \to \infty$). The latter effect has already been discussed in Sec. 7.2. We conclude that the ratio between $\mathrm{Re}(s_\rho)$ and $\mathrm{Im}(s_\rho)$, which are responsible for stretching and absorption, respectively, must be well balanced when approaching the unit cell boundary, such as to achieve a proper attenuation for both evanescent as well as oscillatory tails and keep the errors from limited sampling small.

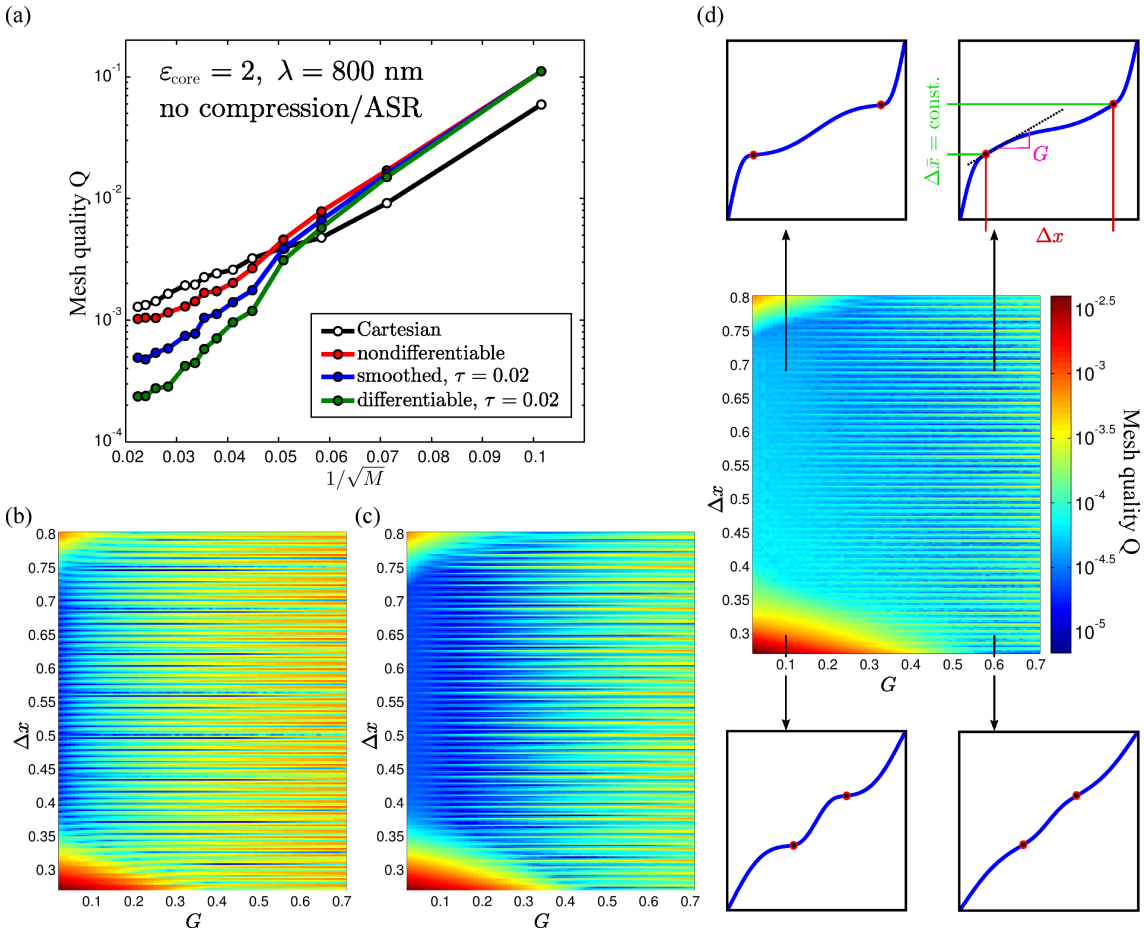


Figure 8.2.: *Analysis of the adapted mesh performance of a circular fiber in a square unit cell. (a) Convergence characteristics of different mesh types without ASR. (b)-(d) Dependence of the mesh quality Q on the ASR parameters Δx and G for a fixed number of plane waves $M = 997$. The same ASR transformation is applied along directions x^1 and x^2. Panel (b) depicts the results of the non-differentiable mesh, (c) of the differentiable one with $\tau = 0.002$, and (d) of the differentiable one with $\tau = 0.015$. The color scale of (d) applies to (b) and (c) as well. Further details see text. Picture adapted from Ref. [100].*

8.2. Performance Comparison of Non-Differentiable and Differentiable Meshes

In order to examine the performance of the non-differentiable, smoothed, and differentiable meshes, we choose an artificial periodic arrangement of step-index fibers of radius $r = 800\,\mathrm{nm}$ within a square unit cell of dimension $a = 4000\,\mathrm{nm} = 5r$ [100]. Considering the errors plotted in Fig. 8.1, this unit cell choice is near the optimum but still large enough to avoid cross talk with neighboring unit cells even for modes close to the cutoff. The core permittivity is chosen as $\varepsilon_{\mathrm{core}} = 2$, the background material is made of air $\varepsilon_{\mathrm{bg}} = 1$. A separate cladding is not considered. Instead, the background material takes its role in the calculation of the analytic reference solutions. The system is analyzed for a wavelength of $\lambda = 800\,\mathrm{nm}$ and the k-space truncation is circular.

The measure of the mesh quality Q is quantified as the maximum relative error of the first ten guided eigenmodes

$$Q = \max(\mathrm{err}_{\mathrm{rel}}(\mathrm{n}_{\mathrm{eff},i})), \qquad i = 1, \ldots, 10\,. \tag{8.3}$$

This means, the smaller Q, the better the mesh. The convergence behavior of Q for all three AC mesh types (without ASR) and an ordinary Cartesian mesh is plotted in Fig. 8.2(a). Without ASR the smoothed and in particular the differentiable mesh performs better than the non-differentiable mesh. All adapted meshes lead to better results than the Cartesian mesh above a certain number of plane waves ($M \approx 220$) which seems to be the minimum number of degrees of freedom required to represent the larger effective permittivities of the adapted meshes properly. For growing M the improvement gets larger. As expected, the results of Fig. 8.2(a) confirm that the differentiable mesh works better than the smoothed mesh, which is what our other empirical studies confirm as well. Therefore, we restrict the further discussion to non-differentiable and differentiable meshes.

The mesh performance comparison with ASR for the same system is shown in Fig. 8.2(b)-(d). Plotted is Q in a pseudo-color plot for a fixed number of plane waves $M = 997$ against the ASR parameters Δx and G. Since the unit cell is square, we apply the same ASR transformation along directions x^1 and x^2. The parameter G has been stepped in 0.01 intervals, and the parameter Δx in 0.0025 intervals. The small drawings above and below panel (d) sketch the characteristics of the ASR transformation function in the indicated parameter space regions.

From the plots we infer that the non-differentiable mesh performance for small values of G, which means a strong compression, is to a great extent independent of Δx and very well — Q is roughly two orders of magnitude smaller than without ASR. However, the best performance is achieved for larger values of G and specific values of Δx. There, the quality is very sensitive to small changes in Δx, and small changes thereof easily result in errors two orders of magnitude larger. A similar behavior is visible in the results for the differentiable mesh, but there the results are clearly less sensitive to small changes in Δx.

We think it is quite interesting to have a look at what ASR does to the effective permittivity. As can be seen in Fig. 8.3 for the non-differentiable mesh, and in Fig. 8.4 for the differentiable mesh, the effective permittivity changes quite dramatically and in a rather unintuitive way. The ASR compression increases in general the maximal values of the effective permittivity components. While component ε^{12} (central row) geometrically remains by and large unaltered, the geometric features of components ε^{11} and ε^{33} become "more periodic". In contrast to the intuitive explanations we found for the effective permittivity with AC (cf. Sec. 7.1.1), the reason for the performance improvement of

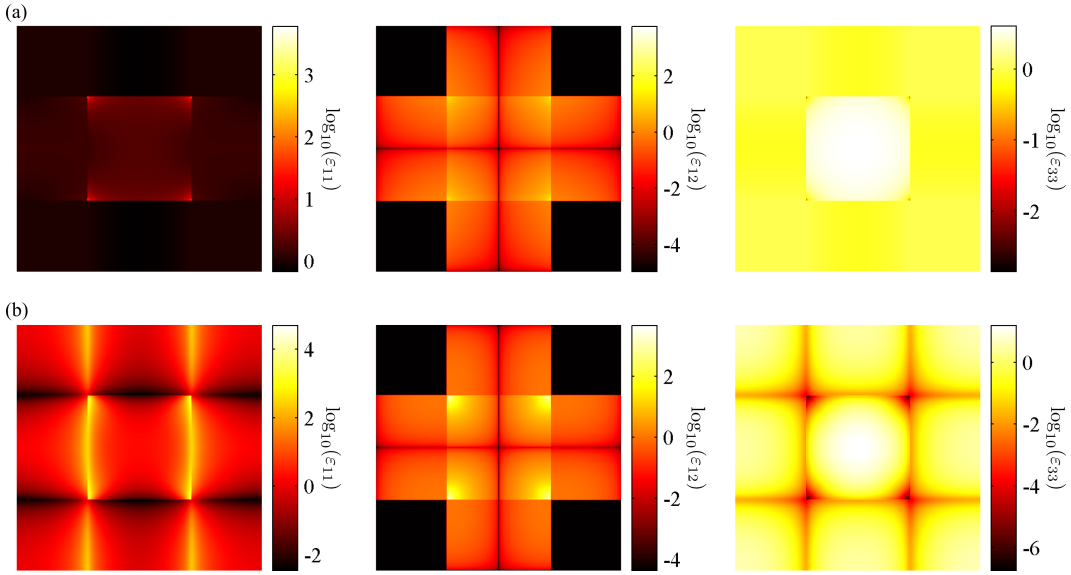


Figure 8.3.: *Effective permittivity of the circular structure with non-differentiable mesh in a log-scale pseudo-color plot: (a) without ASR (cf. Fig. 7.2(c)), (b) with ASR ($G = 0.01$, $\Delta x = \Delta\bar{x} = 0.6/\sqrt{2}$).*

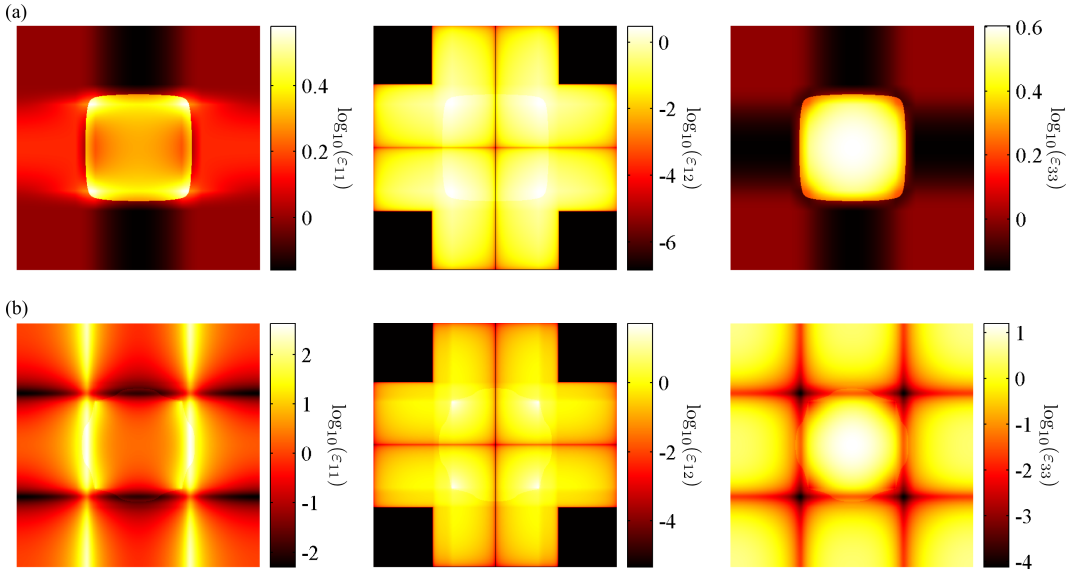


Figure 8.4.: *Effective permittivity of the circular structure with differentiable mesh in a log-scale pseudo-color plot: (a) without ASR (cf. Fig. 7.10(c)), (b) with ASR ($G = 0.01$, $\tau = 0.02$, $\Delta x = \Delta\bar{x} = 0.6/\sqrt{2}$).*

AC with ASR in comparison to plain AC is not intuitive and remains to be clarified in future works.

Since the characteristic oscillatory behavior with changing parameter Δx apparent in Fig. 8.2 remains for other material parameters, real-space discretizations, and degrees of freedom as well [100], we repeat the parameter scans with a much larger resolution to get on its track. Figure 8.5(a) depicts this high resolution scan in a close-up. Striking are the regular error jumps for certain values of Δx from a positive to a negative deviation from the reference values. In between the jumps, Q varies almost linearly. The minimum error is always in the middle between subsequent jump locations. At the same time the error dependence on G is of minor importance, Q barely varies with respect to G. We recall that Δx determines the number of coordinate lines within the structure, and G is related to their density at the inflection points and, thus, the material surface. This finding suggests a connection to the geometrical size of the effective structure. Hence, in Fig. 8.5(b) we depict the relative errors of the individual modes along a line cut at $G = 0.165$ together with the distance between the physical structure surface and the numerical structure surface. For an illustration of these two surfaces see Fig. 8.5(c).

The numerical surface we define as the center between the two sampling points closest to the physical surface. The reason for this choice is that the discretized permittivity has the jump somewhere in between these two coordinates — one sampling point is outside the structure and "sees" the background material, the other is within the structure and "sees" the core permittivity. Let us reconsider the curve progression of reconstructed truncated Fourier series at jump discontinuities (cf. Sec. 3.4.3). They are symmetric with respect to the points horizontally in the center between the sampling points next to the jump, and vertically at half the jump height. In this context, our choice seems well justified. The numerical interface represents the effective size of the structure in truncated Fourier space and, therefore, influences the effective refractive index of the simulated mode.

It is well observable that the distance between the two interfaces exhibits the same sawtooth behavior as the error. This sawtooth stems from the fact that if Δx is varied, a coordinate line eventually crosses the physical structure interface. As a consequence, the numerical interface jumps for about the sampling spacing Δ_ρ (cf. Sec. 3.4.2). This sampling spacing is in turn influenced by the slope G. If G gets smaller, the coordinate lines compression increases, and the sampling spacing decreases. This effect can be nicely seen in Fig. 8.5(d), where the distance is plotted in the same parameter space as in Fig. 8.5(a). For small G the magnitude of the jumps is smaller than for large values of G. Noticeable is also a small asymmetry in the color scale and a slight shift of the null position from the center between the jump positions. This can be attributed to the fact that the slope of the ASR transformation function left and right of the inflection points is often asymmetric. The slope is connected to the coordinate line density, and the density is, in turn, connected to the exact position of the coordinate lines. Hence, the positions are slightly asymmetric as well and so is the distance of the central line to the physical interface.

When the distance between numerical and physical interface is close to zero, the error is smallest because the truncated Fourier series describes a structure with the same effective size as the physical structure which is basis for the analytical reference solution.

In Fig. 8.5(e) we finally plot the individual errors of the guided eigenmodes using the differentiable mesh. The qualitative behavior is very similar as in panel (b). However, the errors do not exhibit the exact sawtooth form as before. An explanation for this change is found when we reconsider that the coordinate lines of the differentiable mesh do not run parallel to the material surface as the coordinate lines of the non-differentiable mesh. This is on the one hand due to the parabola smoothing, and on

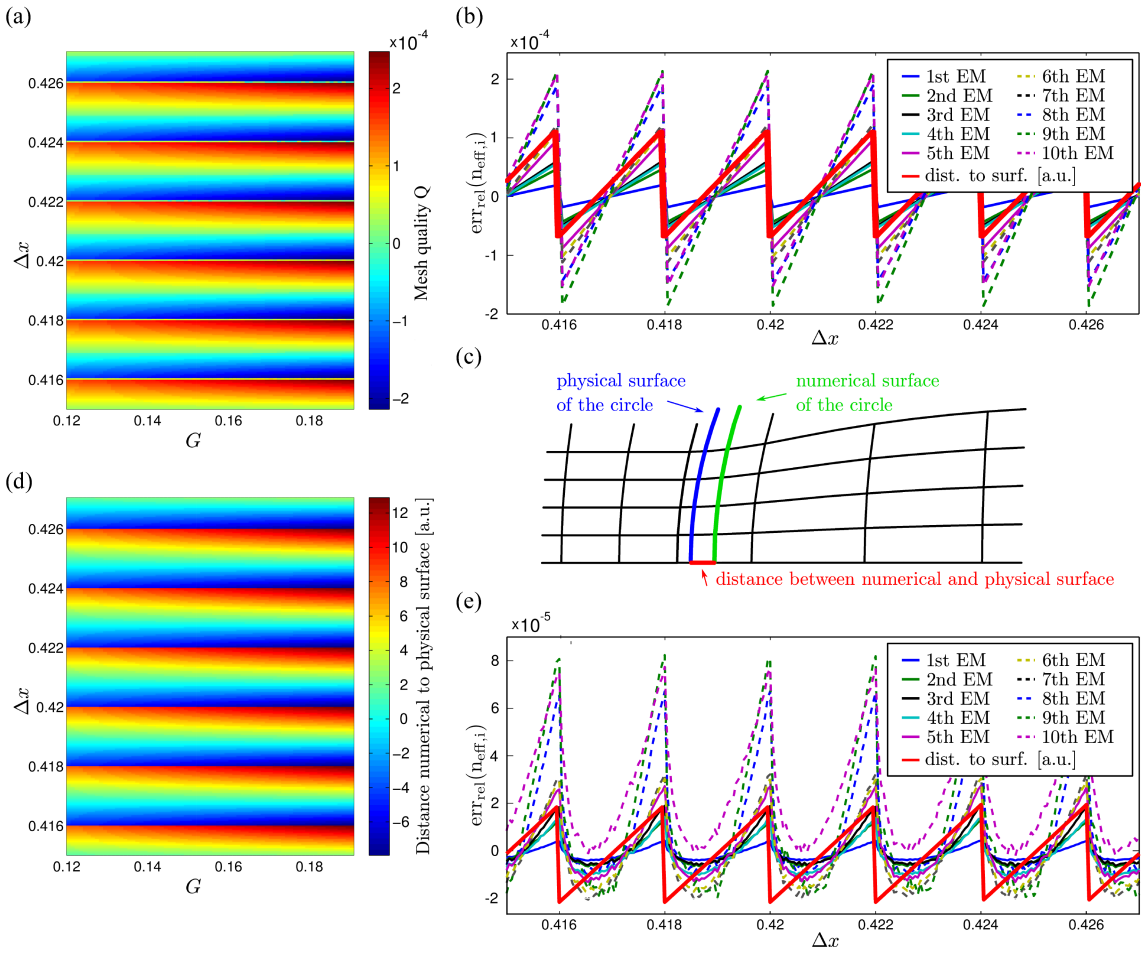


Figure 8.5.: *Examination of the error oscillations along Δx for non-differentiable and differentiable meshes. (a) Close-up of the ASR parameter scan with much higher resolution (steps of 0.001 for G, steps of 0.00005 for Δx) for the artificial periodic fiber system on the non-differentiable mesh ($N_{\mathrm{fft}} = 1000$, $M = 997$). (b) Line cut through (a) at $G = 0.165$. Plotted are the individual relative errors of the first ten guided eigenmodes. (c) Illustration of the numerical structure surface as center line between the two coordinate lines closest to the physical structure surface. (d) Distance between numerical and physical surface in the same parameter range as in (a). (e) Line cut as in (b) for the differentiable mesh ($\tau = 0.002$). Picture adapted from Ref. [100].*

the other hand due to the fact that the enforced boundary conditions in region Ⓑ (cf. Fig. 7.7 and Eqs. (7.21)) do not include an exact (parallel) representation of the physical surface. However, this inexact representation leads to much wider parameter regions where the deviation from the exact physical systems is not nearly zero, but neither as large as with the non-differentiable mesh.

In summary, we observe that the non-differentiable mesh is capable of representing the physical structure very accurately. Its representation can be considerably better than what the differentiable mesh can achieve. At the same time, Q is very sensitive for these meshes and the parameters have

to be chosen very carefully. In contrast, the differentiable mesh is much more robust to the ASR parameter choice and provides almost as good results as the differentiable mesh for a wide parameter range.

For further details, especially for results of systems with larger dielectric constants or metallic cores, see Ref. [100]. In the next section we use the gathered findings to construct optimized meshes for two concrete fiber systems. Furthermore, we evaluate the convergence behavior of these highly optimized meshes with and without PMLs.

8.3. Guided Eigenmodes of an Isotropic Step-Index Fiber

8.3.1. System Setup

The model system we use in the following is a circular step-index fiber with core permittivity $\varepsilon_{\text{core}} = 2.5$ ($n_{\text{core}} \approx 1.581139$), and a silica ($SiO_2$) cladding with $\varepsilon_{\text{clad}} = 2.0952074$ ($n_{\text{clad}} \approx 1.447483$). The setup is schematically depicted in Fig. 8.6(a). The core value is chosen with respect to later applications. The cladding value is determined according to the Sellmeier formula [113], which is a fitting formula for measured material parameters. We use Sellmeier coefficients

$$(B_1, B_2, B_3) = (0.4079426, 0.6961663, 0.8974794)\,,$$
$$(C_1, C_2, C_3) = (0.0684043, 0.1162414, 9.896161)\,\mu\text{m}^2\,,$$

at $\lambda = 1.250\,\mu\text{m}$. The core radius is $r = 2.15\,\mu\text{m}$ and the cladding is (at first) assumed to extend to infinity.

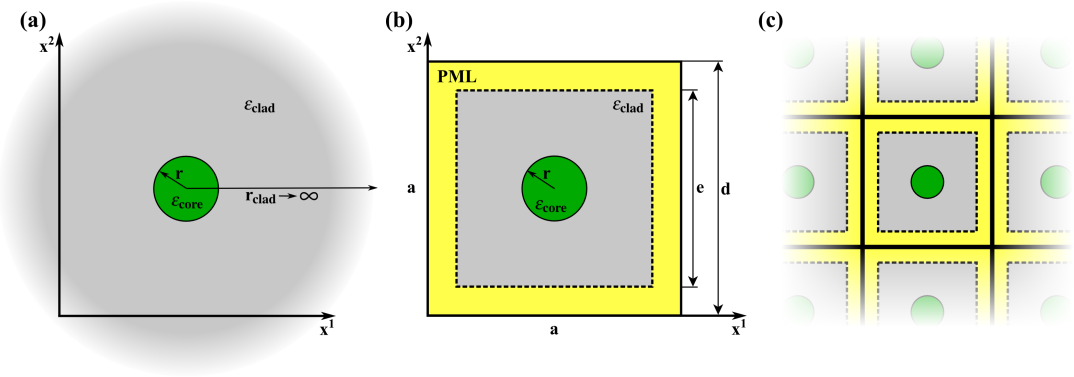


Figure 8.6.: *Schematic illustration of the investigated system. The light propagation is along the z–axis of the depicted right-handed Cartesian coordinate system. (a) We simulate a step-index fiber with isotropic homogeneous circular core (green) and an infinitely extended cladding (gray). (b) For the numerical treatment, the infinite cladding system is squeezed into a finite quadratic domain of edge length a with the help of streched coordinate PMLs (yellow). Thereby, the domain's inner part of size e remains unaltered. (c) Artificial periodic repetition of the unit cells in a quadratic lattice in the transversal plane necessary for the FMM treatment.*

8.3.2. Analytical Eigenmodes and General Numerical Eigenmode Properties

The analytical solutions for the investigated system are obtained with the method described in Sec. 4.1.1 and tabulated in Tab. 4.3. There are in total 24 guided eigenmodes, where ten of them are degenerate. Since we do not consider the degenerate modes, this leaves us with 14 distinct analytic eigenmodes.

The model system used in the FMM is required to meet the periodicity constraints of the method. Because the fiber system does not fit to this criterion, we simply introduce an artificial periodicity as before (cf. Fig. 8.6(c)). Hence, the physical domain of interest is appropriately truncated to form a finite quadratic unit cell, which is periodically continued along the transversal axes of the system. This is of course a tremendous alteration compared to the original problem. In particular, it leads to a fundamental difference between the physical (analytical) and numerical setup. Where the former is rotationally invariant, the latter is treated in a quadratic periodic lattice with a quadratic unit cell. Consequently, the numerical system breaks the symmetry of the problem and lowers the symmetry of the solutions accordingly. For some HE and EH hybrid modes this leads to a lifting of the degeneracy for symmetry reasons: One mode "sees" a certain permittivity environment, where the other "sees" the same but slightly rotated environment. Analytically, the modes span a two-dimensional subspace and the corresponding orthogonal linearly independent solutions differ by a rotation of $\pi/2n$. The numerical solutions are not necessarily orthogonalized, though. The splitting makes those modes much harder to relate to an analytic mode than for the degenerate modes. While HE and EH modes always appear twice in the eigenmode spectrum, the TE and TM modes appear only once because of their rotationally invariant nature mentioned before. Despite the small but inevitable degeneracy lifting, fairly accurate results for guided eigenmodes can be achieved using the periodic FMM, as we have already demonstrated in Sec. 8.2, where quality factors Q below 10^{-5} could be achieved (cf. Fig. 8.2(b)-(d)).

8.3.3. Perfectly Matched Layers

Periodic boundary conditions are in general adequate for the determination of guided eigenmodes of a waveguide, because guided mode fields decay evanescently in the cladding and this decay is usually fast. However, as soon as we introduce a scatterer, the unit cell needs isolation to inhibit energy transfer to its neighbors. This is achieved by the SC-PMLs introduced in Sec. 7.2. Propagating waves impinging onto the PML will thus be attenuated by several orders of magnitude and evanescent tails of guided modes in principle decay even faster in the PML region.

The PML parameters used in the subsequent calculations are: Outer dimension of the PML region $d = 1.001a$, size of the inner boundary of the PML region $e = 0.8a$, and $\gamma = 0.5 + 0.5\mathrm{i}$ (cf. Fig. 8.6(b)). Note that we pick d slightly larger than a which is a necessity of our implementation to avoid divisions by zero in the real-space Π matrix of Eq. (7.39) and Eq. (7.40).

In the following, we investigate two interesting cases. First, we continue the analysis of the fiber with an infinitely extended cladding in Sec. 8.3.4. The second case, discussed in Sec. 8.3.5, is a finite cladding system where we introduce an additional cladding/air interface at radius r_{clad}.

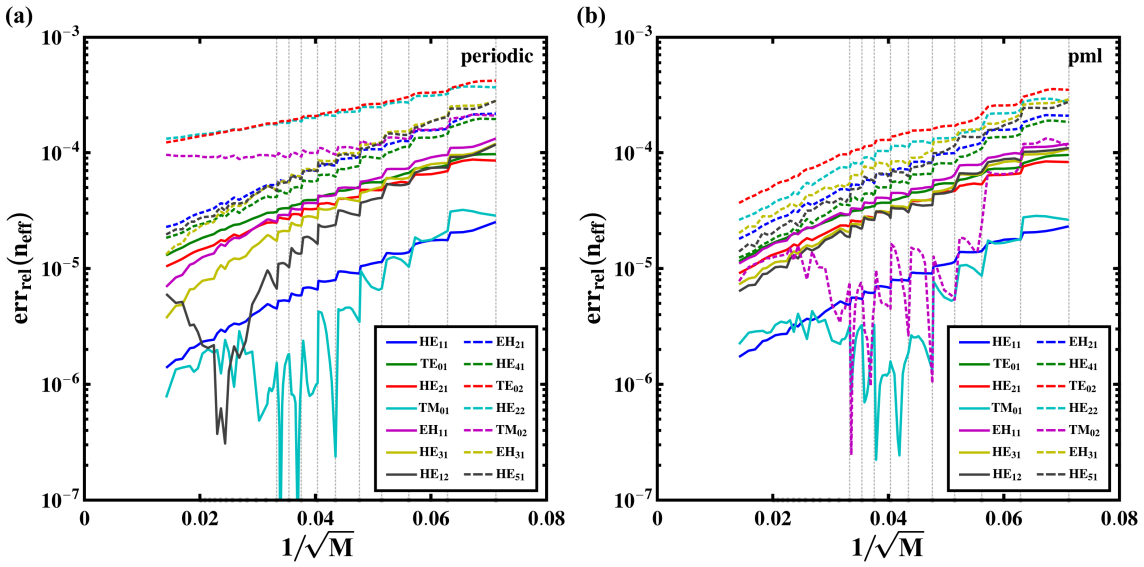


Figure 8.7.: *Convergence plot of the error of the effective refractive index* $\mathrm{n_{eff}}$ *for all 14 guided modes of the infinitely extended cladding system with* $a = 7\ \mu\mathrm{m}$ *and* $\lambda = 1.25\ \mu\mathrm{m}$. *(a) Results for standard FMM (Cartesian mesh) without PMLs, and (b) with PMLs.*

8.3.4. The Infinite Cladding System

We start with the analysis of the fiber with the infinite cladding system (cf. Fig. 8.6(b)). For this system Fig. 8.1(b) suggests a small lattice constant which we choose to be $a = 7\ \mu\mathrm{m} \approx 3.25r$.

Figure 8.7 depicts the convergence behavior of (a) the periodic system without PMLs, and (b) of the PML-isolated system for the number of plane waves varying from $M = 197$ (rightmost) to $M = 4997$ (leftmost). The guided modes converge smoothly towards the analytic reference value, usually approaching from lower values. However, for increasing M, some modes cross the reference value, subsequently approaching the reference value from above. This overshooting appears as broad dips that bring a deviation from the otherwise almost linear curve progression. These dips are visible for example in the HE_{12} curve in Fig. 8.7(a). The little pronounced jumps at certain values of M, where additional reciprocal lattice vectors lead to a noticeable redistribution of energy between the different plane waves, are noteworthy as well. The jump positions coincide with those values of M where our circular truncation scheme in k-space provides a new order of k-vectors on the k_x– and k_y–axes. We mark the respective values of M with grey vertical dashed lines. This observation suggests that plane waves with k-vectors parallel to the lattice vectors of the quadratic real-space lattice contribute more than the others. We will encounter jumps at those positions in almost all subsequent convergence plots again. An explanation of this behavior is given in Sec. 8.4.2. With the rather small lattice constant, we see that in Fig. 8.7(a) the TE_{02}, HE_{22}, and TM_{02} modes are influenced by the missing isolation of the unit cell and converge considerably slower or even towards a different value in comparison to the PML case shown in Fig. 8.7(b). We could avoid this by choosing a larger lattice constant such that this effect would become smaller and would be hidden by other errors. However, with PMLs the isolation prevents cross talk between modes of neighboring lattice sites, and the small lattice constant leads to smaller deviations from the analytic reference

solution as discussed before.

We additionally would like to benefit from AC and ASR meshes. For our purpose we use a non-differentiable mesh, since this type of mesh is expected to provide the best performance for the investigated systems. As discussed above, this is only true if we carefully choose the mesh parameters, which is the topic of the subsequent paragraph.

The mesh construction parameters for the AC are fixed by the system settings. However, the parameters for the ASR need to be optimized; namely, the inclination parameter G, and the parameter Δx which determines the fraction of coordinate lines that resolve the core region. The maximum of the relative errors of all 14 guided eigenmodes' effective refractive indices, Q, for $M = 997$ plane waves is depicted in Fig. 8.8. The maximal error Q is color-coded on a logarithmic scale over a large set of parameter combinations G and Δx. As depicted in Fig. 8.8(a), we first use a rough parameter space resolution (G varying in steps of 0.01, and Δx in steps of 0.01) in order to get an overview and approximately determine the region with the lowest maximum error. In a second step we scan this region with a much higher resolution (stepwidth 0.001 in G and 0.00005 in Δx direction) to find the best parameter set. Figure 8.8(b) shows the part of the high resolution scan that is marked with a white box in Fig. 8.8(a). We only scan a part of the entire parameter space because with the required high resolution the calculations become expensive in total. Thus, this procedure does not necessarily provide the global minimum. However, from our investigations we expect to hit a local minimum with a maximal error close to the global one, because all local minima we find show very similar residual errors. In this case we find $G = 0.144$ and $\Delta x = 0.42655$ with $Q = 4.2536 \cdot 10^{-6}$ as the best parameter set, which we will use in the subsequent adapted coordinate calculations for the small

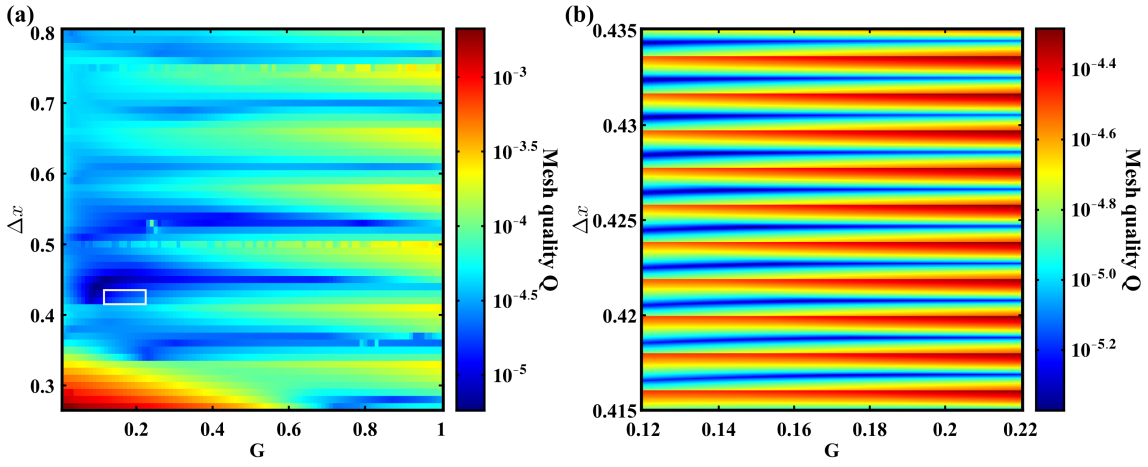


Figure 8.8.: *Parameter scans for the infinite cladding system ($a = 7\,\mu\mathrm{m}$, with PML) in order to find an optimized adapted mesh. Depicted is the maximum of the relative error over all guided modes depending on the input parameters for the ASR — the inclination parameter G and the parameter Δx which determines the fraction of coordinate lines within the core region. (a) shows a coarse scan and (b) a high resolution scan of the white boxed region in (a) containing the local minimum at $G = 0.144$ and $\Delta x = 0.42655$.*

infinite cladding system.[2] The optimized adapted mesh is visualized in Fig. 8.12(a). It is necessary to emphasize that these mesh parameters optimize the adapted mesh only for the specific system at that specific number of plane waves $M = 997$. The plots show that the parameter G is of minor importance to the maximum relative error, because the error is almost constant for a wide range of G (and constant Δx). More crucial is the parameter Δx. Where in the coarse scan a super structure oscillation in Δx with a large period of about 0.08 is visible, the detailed scan shows a much smaller correct period of about 0.002. This oscillation is caused by the discretization of the permittivity as discussed above.

With the obtained optimized mesh we perform a convergence study. Figure 8.9 depicts the relative errors of the guided modes' effective indices. The convergence behavior fundamentally changes in comparison to the convergence behavior obtained with a Cartesian mesh (cf. Fig. 8.7). Where the latter is smooth, the former does not converge uniformly but shows jumps of orders of magnitude especially for small numbers of retained expansion orders. In Fig. 8.9 on the right-hand sides of the

[2] We cover a *large* infinite cladding system with $a = 21\ \mu\text{m}$ later on.

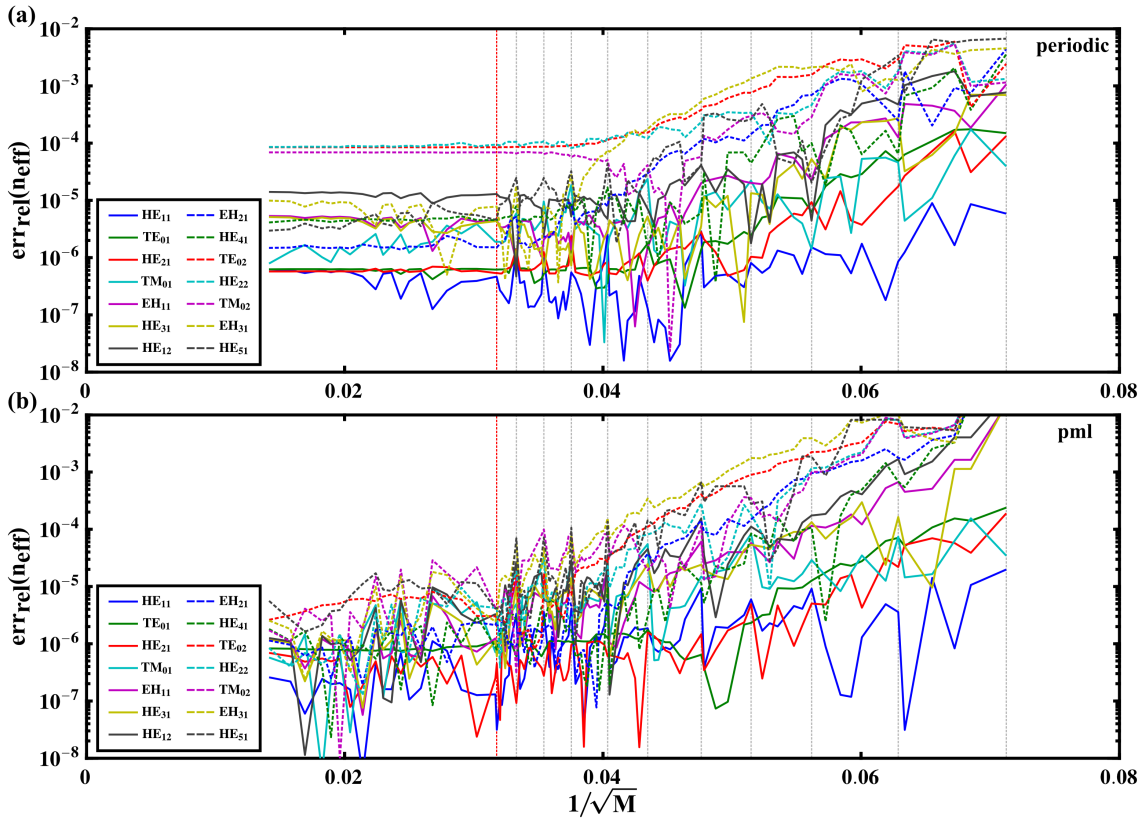


Figure 8.9.: *Convergence behavior for the guided eigenmodes calculated with the optimized adaptive mesh (infinite cladding system). Panel (a) depicts the results without PMLs where the relative errors converge towards a finite offset from the reference values. Panel (b) shows the relative errors for the isolated unit cell. The vertical red dotted line indicates the number of modes used for the mesh optimization — there, the errors have a local minimum.*

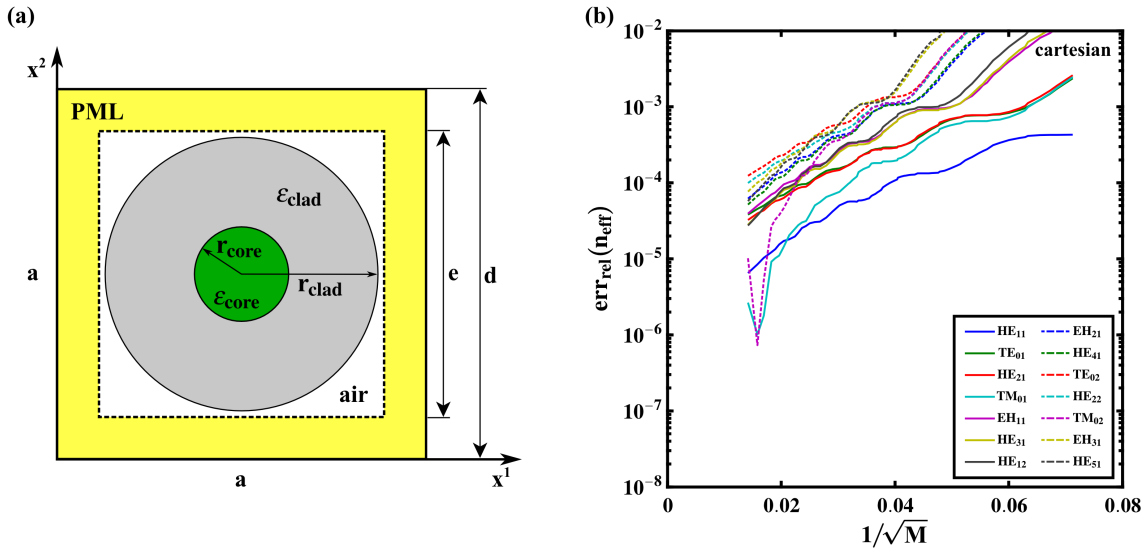


Figure 8.10.: *(a) Schematic illustration of the finite cladding system unit cell. For the numerical calculation we choose* $r = 2.15\,\mu\mathrm{m}$, $r_{\mathrm{clad}} = 7.5\,\mu\mathrm{m}$, *and* $a = 21\,\mu\mathrm{m}$. *(b) Convergence behavior for the guided eigenmodes of the large infinite cladding system with* $a = 21\,\mu\mathrm{m}$ *calculated with the Cartesian mesh (with PML).*

depicted regions, the errors are of the order of 10^{-2} and, therefore, almost one order of magnitude larger than with the Cartesian mesh. But on the left-hand sides, the errors are well below the errors for the ordinary FMM. In the convergence plot of the periodic system shown in Fig. 8.9(a), the effective refractive indices converge towards values that are quite different from the solutions of the open system described by the analytic reference values. However, the results nicely agree with the results obtained from the Cartesian mesh (cf. Fig. 8.7(a)). The influence of the PML isolation becomes apparent in Fig. 8.9(b). Here, the deviation has its minimum at the point where the mesh was designed for, at $M = 997$, which corresponds to $1/\sqrt{M} \approx 0.03167$ (dotted red vertical line). There, the maximal error drops from $Q_{\mathrm{cart}} = 9.6969{\cdot}10^{-5}$ for the Cartesian mesh to $Q_{\mathrm{adapt}} = 4.2536{\cdot}10^{-6}$ which is more than one order of magnitude. Interestingly, with an even increasing number of plane waves the error first increases, but finally decreases again for very large M without significantly outperforming the design point. We conclude that it is advisable to optimize the mesh exactly for the number of plane waves one is going to use for the calculations later on. We have seen that with the appropriately designed adapted mesh, the relative error can be reduced well below 10^{-4} because the mesh helps to correctly resolve the physical interface between core and cladding. The remaining deviation is not smooth anymore, but reveals characteristic spikes of the order 10^{-5} and below. It will be topic of Sec. 8.4.2 to explain where some of theses sharp features come from. For the time being we have a look at the finite cladding system.

8.3.5. The Finite Cladding System

The finite cladding system, schematically illustrated in Fig. 8.10(a), consists of the core and cladding as described before, but with an additional cladding/air interface. Even though realistic fibers often have outer cladding dimensions of $125\,\mu\mathrm{m}$, we think a radius $r_{\mathrm{clad}} = 7.5\,\mu\mathrm{m}$ and a lattice constant

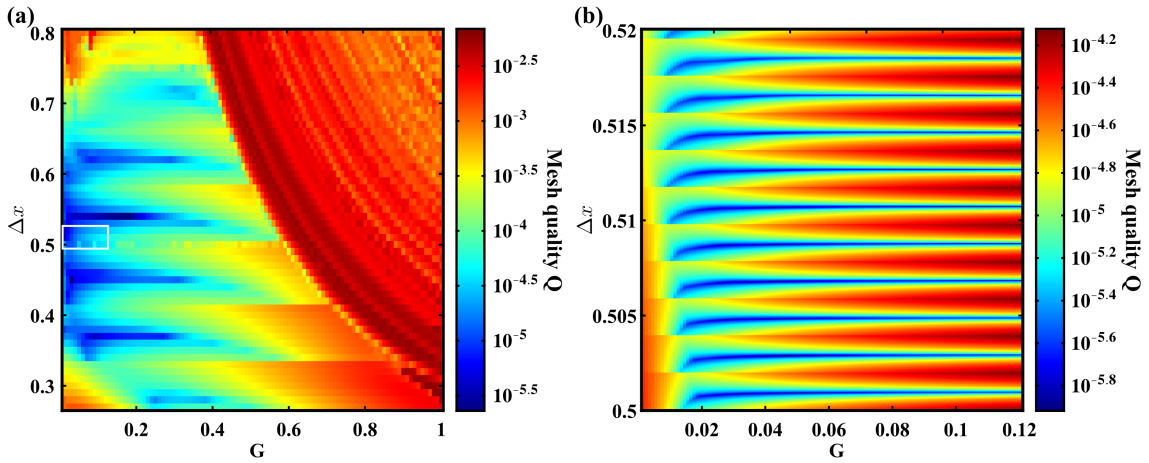


Figure 8.11.: *Mesh optimization parameter scans for the infinite cladding system of size* $a = 21\,\mu\mathrm{m}$ *(with PML). Depicted is the maximum of the relative error* Q *over all guided modes depending on the input parameters for the used ASR. Panel (a) shows a coarse scan, and panel (b) depicts a high resolution scan of the white boxed region in (a) containing the local minimum at* $G = 0.103$ *and* $\Delta x = 0.51070$.

$a = 21\,\mu\mathrm{m} \approx 9.77r$ is a reasonable choice. This choice is motivated by the discussion about the method's convergence behavior for small core size to lattice constant ratios above. Furthermore, we argue that an additional cladding/air interface, sufficiently apart from the core, should have a rather small influence on the guided modes because their evanescent tails quickly decay. We will show in the end of this section that the latter is indeed the case, because the cladding/air interface has an influence on the effective refractive index which is at the most in the order of our numerical errors or below.

Because we change the lattice constant a in comparison to the infinite cladding system while the core radius stays the same, we cannot reuse the optimized mesh of Sec. 8.3.4. Instead, we have to repeat the process of mesh optimization from the beginning. Furthermore, we only have analytic reference solutions for infinite cladding systems at hand. Therefore, we perform the mesh optimization procedure for the large lattice constant ($a = 21\,\mu\mathrm{m}$), but without the additional interface. After that, we show that the optimized mesh for the infinite cladding system also performs well for the finite cladding system with the cladding/air interface.

Please note that the analytic *single cylinder mesh* we use only properly resolves the core/cladding interface. An additional resolution of the cladding/air interface would necessitate a totally different type of analytic mesh — a *double cylinder mesh* for nested cylinders (cf. Ref. [100]) — which can be constructed with the construction schemes from Sec. 7.1.1, but is not rigorously analyzed yet. Therefore, we delay the optimization of such meshes to future work.

The mesh optimization data is shown in Fig. 8.11. Again, we first perform a coarse scan shown in panel (a), and then a fine scan of the parameter region marked with the white box, depicted in panel (b). The resolutions are the same as before. The area on the right-hand side of panel (a), where large errors were obtained, is an artifact of the mesh. There, the parameters G and Δx reach values such that the mesh starts folding, i.e., the coordinate mapping is not bijective anymore. Note

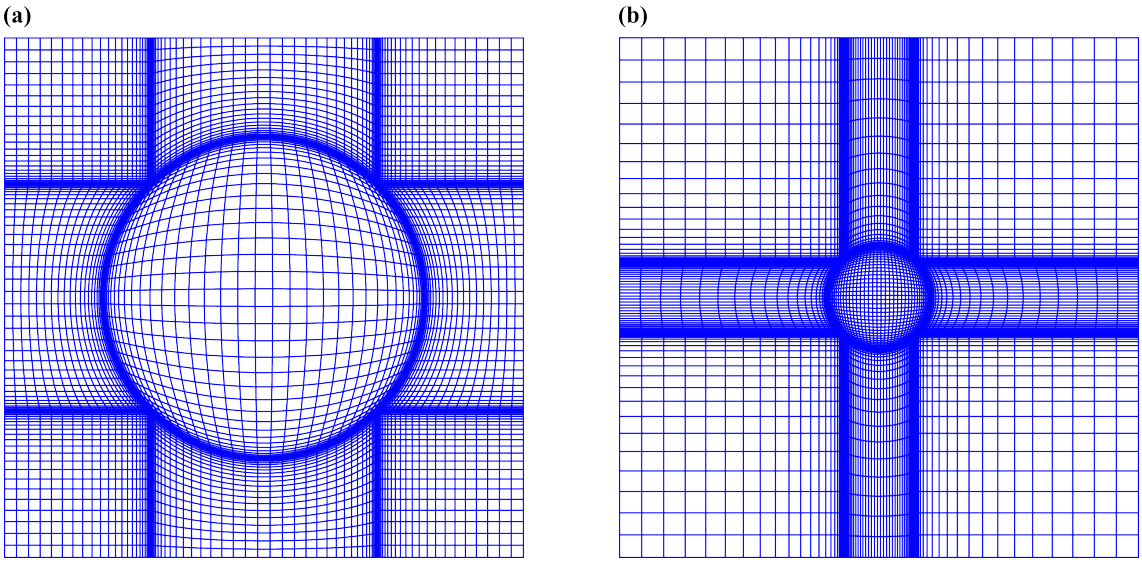


Figure 8.12.: *(a) Optimized adapted mesh of the small infinite cladding system ($a = 7\,\mu\mathrm{m}$). (b) Optimized adapted mesh of the large infinite/finite cladding system ($a = 21\,\mu\mathrm{m}$). Both visualizations show 80 coordinate lines in each direction which are evenly distributed over the whole unit cell in the transformed coordinate space.*

that the coarse scan suggests optimal parameters at around $\Delta x = 0.54$ and $G = 0.2$. This is misleading, because there we just hit one of the horizontal dark blue stripes visible in the fine scan of Fig. 8.11(b) by coincidence. These parameter regions provide good results but are very narrow in Δx. The smallest local minimum in the maximal relative error we could find is at $G = 0.103$ and $\Delta x = 0.51070$. There are, however, many other parameter sets that show quite similar performance in the narrow horizontal stripes described above. The optimized adaptive mesh is visualized in Fig. 8.12(b).

Figure 8.13(a) depicts the convergence behavior of the guided eigenmodes of the large infinite cladding system with an optimized adapted mesh. For the designed number of retained plane waves $M = 997$ (indicated by the vertical red dotted line), the maximal deviation is improved from $Q_{\mathrm{cart}} = 8.1233 \cdot 10^{-4}$ for the simple Cartesian mesh, shown in Fig. 8.10 (b), to $Q_{\mathrm{adapt}} = 1.2098 \cdot 10^{-6}$ for the simulation with the optimized adapted mesh. Compared to the system with the small lattice constant in Sec. 8.3.4, the positive effect of the optimized mesh is even more pronounced.

The influence of the additional cladding/air interface can be estimated by comparing Fig. 8.13(a) with Fig. 8.13(b). In the latter, the error has been calculated relative to the available analytic solutions of the infinite system as well. Qualitatively, the two plots show the same behavior. This means that, first of all, the interface changes the effective indices only slightly. This result is also in accordance with simulations using a Cartesian mesh (not shown). And secondly, the mesh optimized for the infinite cladding system provides a similar performance for the finite cladding system even though the coordinate lines do only properly align with the core/cladding interface. This is not surprising since the important physical effects are dominated by the properties of the core region, which is the same in both cases and well discretized with the specifically tailored adapted mesh.

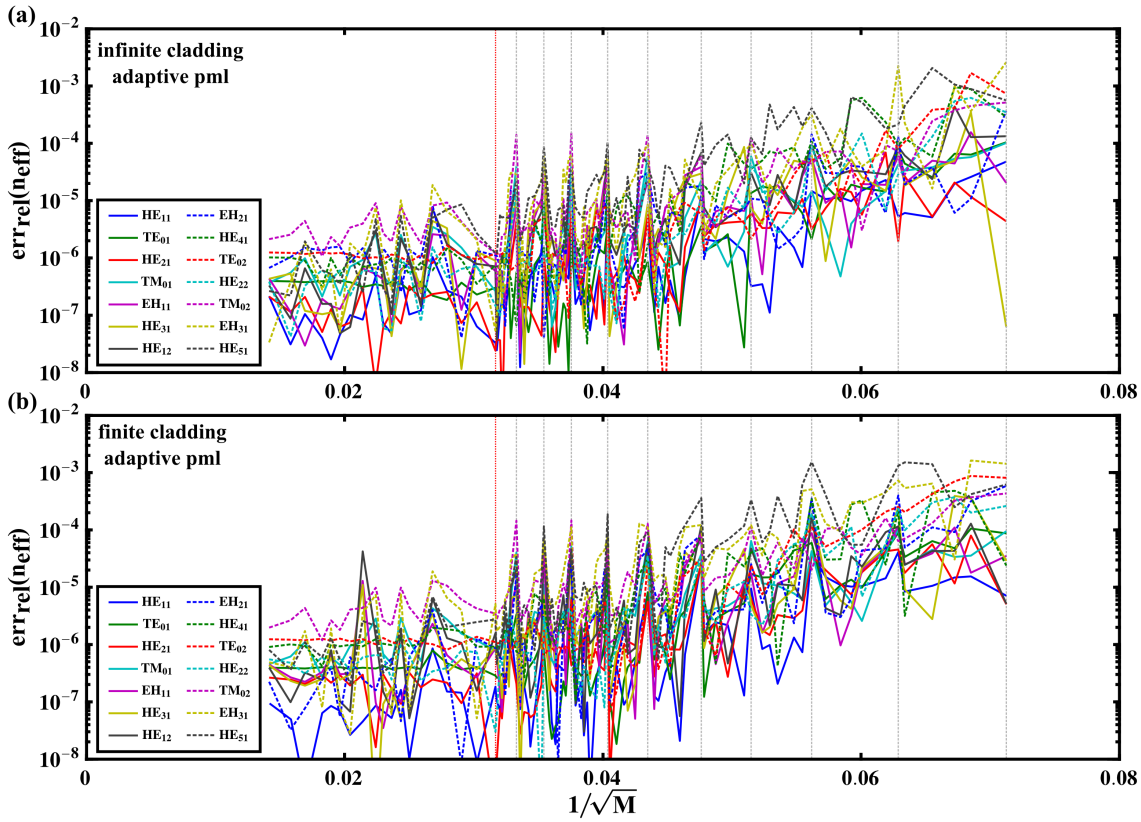


Figure 8.13.: *Convergence behavior for the guided eigenmodes of the system with optimized adaptive mesh ($a = 21\,\mu\mathrm{m}$, with PML) calculated with (a) infinite cladding, and (b) additional cladding/air interface at $r_{\mathrm{clad}} = 7.5\,\mu\mathrm{m}$. The vertical red dotted lines indicate the number of modes used for the mesh optimization.*

8.3.6. Variation of the Core Permittivity

So far, we have investigated an isotropic core. However, the LPG we present in Sec. 9.2 has an anisotropic core. Therefore, we have to argue and show that the optimized mesh still performs well in the anisotropic case. It is hard to prove this directly because we lack an analytic inhomogeneous anisotropic solution we can compare to. The best we can do is to test the mesh for isotropic cores, but vary the core permittivity in the range of the values for the anisotropic material. The limits are the extremal values of the ordinary and extraordinary refractive indices of the used liquid crystals which are wavelength-dependent (cf. Sec. 9.2). The chosen permittivity $\varepsilon_{\mathrm{core}}$ ranges from 2.15 to 2.85 in steps of 0.005, which covers all liquid crystal permittivity values in the desired wavelength range. In addition, to make the test more realistic, we choose silica for the cladding material. Silica is dispersive and its permittivity varies from $\varepsilon_{\mathrm{clad}} = 2.1037107$ at $\lambda = 1\,\mu\mathrm{m}$ to $\varepsilon_{\mathrm{clad}} = 2.0875990$ at $\lambda = 1.48\,\mu\mathrm{m}$ (cf. Sellmeier above). Once again, we use the large infinite cladding system to benefit from appropriate reference solutions.

The results of our calculations are presented in Fig. 8.14. The number of analytically determined guided modes is plotted in panel (a). The maximum relative error for the Cartesian mesh, depicted

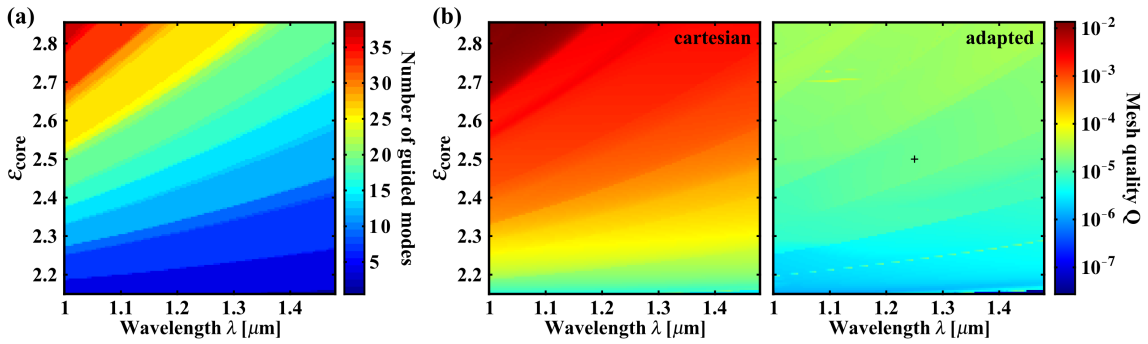


Figure 8.14.: *Properties of the large ($a = 21\,\mu\text{m}$) fiber system depending on the core permittivity $\varepsilon_{\text{core}}$. The core permittivity is varied in the value range of the later used anisotropic material. The silica cladding is dispersive and treated accordingly. (a) Number of analytically determined guided eigenmodes. (b) Comparison of the maximal relative error of all guided modes' effective refractive indices obtained with a Cartesian (left) and the optimized adapted (right) mesh. The black + marks the (material) parameters used in the mesh optimization procedure.*

in panel (b) on the left, is directly connected to the number of guided modes — where new guided modes emerge from below the cutoff, the error usually jumps to higher values. This is partly also true for the results obtained with the optimized adapted mesh, but less pronounced than with the Cartesian mesh. In general, the adapted mesh provides much more accurate results, as can be seen on the right of panel (b). Where the *largest* maximal error for the Cartesian mesh in the whole parameter space is $Q_{\text{cart,max}} = 1.3583 \cdot 10^{-2}$, the same for the optimized mesh is $Q_{\text{adapt,max}} = 5.4617 \cdot 10^{-5}$. Both values are obtained at the upper core permittivity limit. The *smallest* Q values near the lower core permittivity boundary are $Q_{\text{cart,min}} = 2.8449 \cdot 10^{-6}$ and $Q_{\text{adapt,min}} = 2.7856 \cdot 10^{-8}$. It is interesting to observe that the largest errors are obtained for high index contrasts, where most of the modes are tightly confined to the core, but the field patterns of the modes near the guiding cutoff vary on small length scales. This indicates that the limiting factor for accuracy is the small number of plane waves or, correspondingly, the low order of the largest retained k-vectors.

With these simulations we could show that even though the adapted mesh was designed for a parameter set near the center of the investigated region ($\varepsilon_{\text{core}} = 2.5$, $\varepsilon_{\text{clad}} = \varepsilon_{\text{SiO}_2,\lambda=1.25\,\mu\text{m}}$), it provides excellent results for the whole parameter range covered by the used liquid crystal material.

8.4. Issues with Coordinate Transformations

The preceding sections merely discussed the real part of the guided modes' effective refractive indices. We neglected two crucial problems we would like to discuss now: First, the FMM provides as many eigensolutions as plane waves used in the expansion — the challenge is to establish some reliable and robust rules how to automatically find the *guided* modes among them. The second topic concerns the way how coordinate transformations affect the spectrum of the eigenmodes' propagation constants.

In the discussion we distinguish between real coordinate transformations of AC and ASR, and com-

plex coordinate transformations of PMLs.

In a non-absorptive periodic system calculated with standard FMM, e.g., the artificial periodic arrangement of step-index waveguides without PML isolation discussed in the beginning of Sec. 8.3.4 (cf. Fig. 8.7(a)), the system is described by the small eigenvalue equations (cf. Eqs. (6.13)). The diagonalization of the system matrix results in purely real (positive and negative) eigenvalues. Since the eigenvalues are the propagation constants squared (γ^2), this always yields purely real or purely imaginary propagation constants, corresponding to propagating and evanescent eigenmodes, respectively. Even though the operator is not Hermitian per se, we have never observed complex propagation constants. The guided eigenmodes are those modes of the eigenmode spectrum with the highest real effective refractive index values. Thus, the task of guided mode determination is straight-forward for such systems.

8.4.1. Real Coordinate Transformations

When we use adapted meshes with ASR coordinate line compression, the eigenvalue spectrum changes fundamentally. The guided modes are still existent. However, the spectrum additionally comprises complex modes that come in pairs: One eigenmode has the value $v = v_1 + \mathrm{i}v_2$, the other mode the value $v' = v_1 - \mathrm{i}v_2$. Furthermore, additional modes appear with purely real propagation constants within the guided modes effective index region and above. Hence, it is a challenging task to distinguish the truly guided eigenmodes of the waveguide from the spurious numerical modes stemming from the use of adaptive meshes.

In the case of the large infinite cladding system ($a = 21\,\mu\mathrm{m}$) the eigenmode spectrum (real parts) versus number of plane waves is shown in Fig. 8.15(a). Red data points denote guided modes and black dots all other modes which comprise cladding modes below the guiding cutoff (n_{clad}) as well as spurious ASR modes. The latter also occur above the guiding cutoff, where we expect only guided modes, and even above the core refractive index, where no physical modes can live. With increasing M their effective indices continuously decrease. For $M \geq 2785$ they are smaller than the upper core threshold, and finally for $M \geq 3489$ they disappear in the region below the guiding cutoff. In contrast, for the infinite cladding system with the small lattice constant ($a = 7\,\mu\mathrm{m}$), spurious modes are noticeable only for $M \leq 221$ plane waves (not shown).

The main difference between both cases is the amount the adapted mesh is distorted compared to the Cartesian mesh. We find that the smaller the ratio core size over lattice constant, the larger the mesh distortion of the adapted mesh. This distortion translates into increased effective permittivity values as can be seen in Fig. 8.3 and Fig. 8.4. The effective permittivity includes the physical permittivity as well as artificial contributions due to the coordinate transformation and the metric. Even for isotropic permittivities ε this leads to a planar anisotropy. It is to be understood that the same arguments and procedures are applicable to the permeability as well, even though we omit to mention it explicitly here and in the following.

The electric field distribution of the spurious mode with the highest real effective refractive index for $M = 997$, which is marked in Fig. 8.15(a) with the black arrow, is visualized in Fig. 8.15(b). The field pattern is very typical for these spurious ASR modes. The field intensity seems to accumulate in the cladding next to the coordinate lines where the coordinate compression is highest (cf. Fig. 8.12).

We put our observations on record: ASR leads to spurious modes which are, on the one hand, dependent on the concrete real coordinate transformation and, on the other hand, on the number of plane

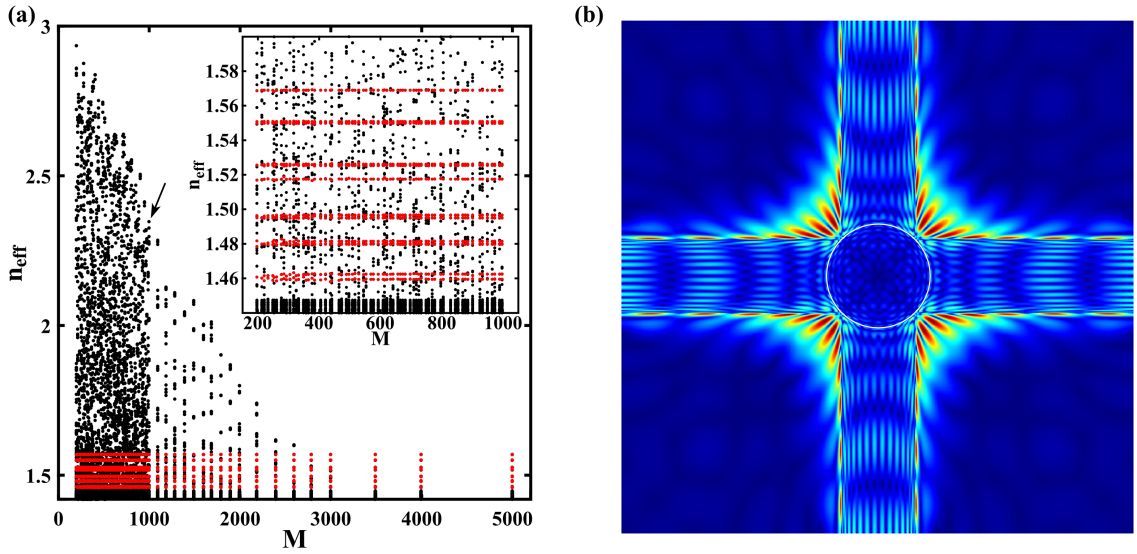


Figure 8.15.: *(a) Convergence plot of the eigenvalue spectrum ($\mathrm{n_{eff}}$) for the large infinite cladding system ($a = 21\,\mu\mathrm{m}$). Shown are results obtained with the optimized mesh, but without PML coordinate transformation. Red dots denote guided modes and black dots all other modes which comprise cladding as well as spurious ASR modes. The inset highlights guided and spurious modes in the guided mode regime. (b) Electric field distribution ($|\mathbf{E}|$) of the spurious mode with the highest real effective refractive index for $M = 997$. The visualized mode is marked in panel (a) with a black arrow.*

waves used in the truncated Fourier series. The larger the distortion (compression) of space, the more spurious modes appear above the guiding cutoff and the higher are their maximal effective refractive indices. Because the modes disappear below the guiding cutoff with an increasing number of retained basis functions, spurious ASR modes seem to be an effect of slow convergence. This suggests the following conclusion: An ASR coordinate line compression helps to accurately represent the structure surface and, thus, the propagation constant of guided modes. But a too strong compression leads to an increasing number of spurious modes with effective indices within and above the guided mode region.

8.4.2. Complex Coordinate Transformations

The improved accuracy with PMLs demonstrated above does not come without complications, either. If only Cartesian meshes and isotropic materials are required, the PML can be implemented as a 3D generalization of Lalanne's proposal [99]. This equation-transform k-space strategy implementation of the eigenproblem is given in App. B.4. The complex coordinate shifting introduces imaginary parts in the permittivity, providing a non-Hermitian, non-symmetric, complex eigenoperator. Consequently, additional PMLs *always* result in complex propagation constants. Even the guided modes gain imaginary parts that are at best in the order of the error of the real part, but are usually larger due to the overlap of their evanescent tails with the lossy PML region. Nevertheless, they can by and large be distinguished by the value of their real parts above the guiding cutoff and at the same time

comparatively small imaginary parts.

We also implemented and tested the PMLs with the structure-transform real-space strategy which is integrable with AC and ASR. We have motivated in Sec. 7.4.3 that even though both strategies are completely equivalent on an analytical level, they differ in numerical details. We expected better results from the real-space strategy, since the effective permittivity can be computed analytically in real-space before the Fourier transformation.[3] Hence, the errors of successive convolutions of truncated Fourier series do not add up.

Similarly to the AC transformations, the *structure-transform real-space* implementation of PMLs introduces spurious modes that contaminate the eigenvalue spectrum, also in the range where one would expect guided modes only. One of our observations is that these spurious modes also influence the truly guided modes: They tend to "push" the guided modes' propagation constants a little away from the optimal values if their propagation constants are similar to the guided modes' propagation constants.[4] Thus, considerable effort has to be put into the determination and selection of the correct guided modes.

Similar to the real coordinate transformations above, our SC-PMLs involve compressions of the mesh. However, the compressions are usually much larger as we try to squeeze the infinite space (or at least a large part of it) into the small PML layers of the unit cell. Thus, a similarly slow convergence effect could be the reason for the spurious modes, but with the available number of plane waves we could not observe a comparable vanishing of the spurious modes below the guided mode cutoff. We guess with our limited number of available degrees of freedom we are still far away from the threshold. The limitations are mainly availability of memory and computational costs for the system matrix diagonalization.

The most difficult challenge with eigenmodes in aperiodic systems is connected to the observation that occasionally the imaginary part of some guided (and continuum) modes flips sign. By this we mean that for example the imaginary part β_I of a forward guided mode (with positive real part of the propagation constant β_R) gets negative, and the corresponding phase factor

$$e^{\mathrm{i}(\,\beta_\mathrm{R}\,-\,\mathrm{i}\,|\beta_\mathrm{I}|\,)\,x^3} \;=\; e^{\mathrm{i}\,\beta_\mathrm{R}\,x^3}\;\cdot\; e^{+|\beta_\mathrm{I}|\,x^3} \tag{8.4}$$

grows in amplitude. This literally means that this mode *gains* energy, which is of course contradictory to a passive or lossy PML medium and not a physically reasonable behavior. Unfortunately, we cannot predict which modes are affected because it does not occur in a regular pattern. In different constellations (wavelength, lattice constant, geometry, permittivities, etc.) different modes are affected. Of course, we *carefully* checked that it is not a random effect due to a sloppy implementation — the sign flips are deterministic events. Furthermore, in the large eigenvalue problem, the corresponding backward propagating solutions are always equivalent (with overall negative sign for the propagation constant, of course). We conclude that there must be a correspondence to the way the problem is formulated.

An example is shown in Fig. 8.16(a), where the eigenvalues of the system are plotted in the complex plane for $M = 997$. Since this numeric effect noticeably disturbs the physical meaning of the solutions and leads to transmittance values above one, we interfere with the calculation in these cases and manually flip signs back to obtain slightly lossy modes instead. The corresponding eigenvalues are indicated by red crosses. We believe that this (drastic) measure is justified on the basis of strong

[3] Except for automatically constructed meshes.

[4] This "pushing" appears to be similar to avoided crossings of bands.

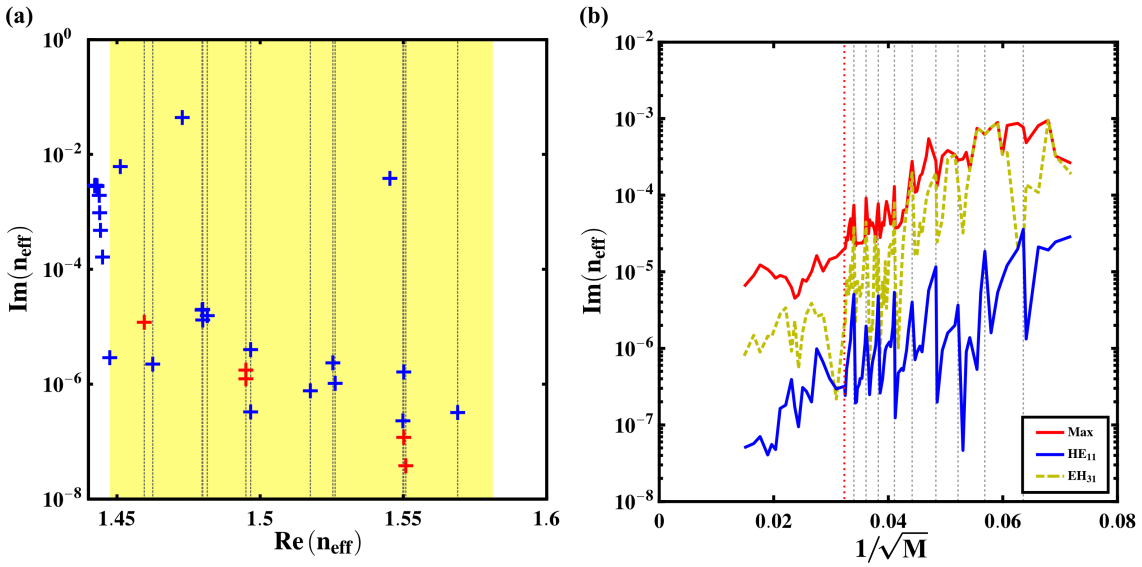


Figure 8.16.: *(a) Eigenvalue spectrum in the complex plane for the infinitely extended cladding system with $a = 7\,\mu\mathrm{m}$ calculated with the optimized adapted mesh and PMLs. The yellow shaded area is the guiding region between lower cutoff and upper core threshold. Blue crosses describe the eigenvalues, red crosses describe the modes where we manually flipped the sign of the imaginary part. Vertical dashed lines indicate the values of the analytical solutions. (b) Convergence plot of exemplarily selected guided eigenmodes' imaginary parts (absolute values) for the same system. The red line marks the maximum value over all guided modes.*

evidence that the eigenmodes with mixed-sign propagation constants are purely due to a erroneous numerical representation of the PML coordinate transformation with truncated Fourier series.

To adduce this evidence, we recall the discussion of Sec. 7.4. Within the equation-transform k-space strategy, we calculate and use the Fourier representations of the bound functions f_ρ. Analytically, the imaginary part in the inner physical domain ($[-e/2, e/2]$) should be exactly zero since it represents the free space. However, the imaginary part of the respective truncated Fourier series (cf. Fig. 7.16) exhibits small oscillations ($\sim 10^{-3}$) around zero within this region. Positive values represent gain and negative values represent loss. Imagining the overlap of eigenmodes with the numerically represented structure, it seems reasonable to assume that in certain constellations when the mode is mainly confined to a small region in the center, the gain could overcompensate the strong attenuation within the PMLs. We actually never observed this situation and believe that this is due to the relatively small magnitude of the oscillations. Nevertheless, it could also be a hint why even the tightly confined lowest guided modes exhibit rather large imaginary parts of the effective refractive indices which only slowly and non-smoothly converge, as can be exemplarily observed in Fig. 8.16(b). Unfortunately, the residual oscillations in the truncated Fourier series of f_ρ seem to be unavoidable due to the chosen basis. With an optimized and specifically designed coordinate transformation it might be possible to decrease the effect, though.

The situation gets far worse if we consider the structure-transform real-space strategy which we have employed for our simulations including AC and ASR where the mixed sign propagation constants

(sign flips) occurred. Instead of the bounded functions f_ρ, the Fourier transformation involves the entries of PML operator $\underline{\underline{\mathbf{\Pi}}}$ introduced in Eq. (7.39) and Eq. (7.40). Its matrix entries additionally comprise the inverse functions $1/f_\rho = s_\rho$ which represent the slope in the transformation functions F_ρ. This slope can become considerably steep, resulting in large values for s_ρ. This is particularly the case for Lalanne's version where s_ρ grows to infinity when $x^\rho \to \pm d/2$. In order to avoid a division by zero, which is numerically not representable in real-space, we pick d slightly larger than one.[5] This trick can also be used to limit the values of s_ρ in Lalanne's formulation. Still, due to the general tangent-like form, when moving from $e/2$ to $d/2$, the slope is for a wide range rather small but then increases very rapidly. This means an increase of d by only one percent ($d = 1.01$) reduces the covered physical space from infinity to about 2.7 times the unit cell size ($e = 0.8$). In this case, the maximal slope of the imaginary as well as the real part is still about 180. The imaginary part of the corresponding periodic function, thus, features a very small and high peak.

The CFS formulation has the advantage that we can adapt s_ρ to our needs directly. For a similar scenario with roughly the same effective size of the considered domain in physical space, the parameters are $\kappa_{\max} = 40$, $\sigma_{\max} = 36$, and $m_\kappa = m_\sigma = 3$ ($a_{\max} = 0$, $m_a = 1$, $\omega = 1$, $e = 0.8$). Due to the third order polynomial grading, the corresponding peak in the imaginary part of s_ρ is considerably lower and broader.

These observations manifest themselves in the k-space representation of the components of $\underline{\underline{\mathbf{\Pi}}}$, which is basically the PML transformed permittivity of free space. The imaginary part of the PML transformed structure's effective permittivity is dominated by their behavior. They are depicted in Fig. 8.17. The $\tilde{\Pi}^{11}$ component is shown in panel (a), on the left side for Lalanne's formulation and on the right for the CFS formulation (parameters as above). Depicted in pseudo-color plots on a logarithmic scale are the absolute values of the Fourier coefficients' imaginary parts over the Fourier orders in $\mathbf{k}^1$ and $\mathbf{k}^2$ direction. The top row shows the complete set of coefficients ($N_{\text{fft}} = 1024$), the bottom row a closeup of the coefficients up to a truncation order of $|m_\rho| \leq 20$ in each (transverse) k-space direction ($M = 1681$). The latter are those coefficients we will take into account for a real-space reconstruction below. We can observe that in the vertical ($\mathbf{k}^2$-) direction on the central line ($m_1 = 0$) the coefficients decrease very slowly. This is due to the function s_y which has its peak in this direction. The coefficients from the CFS function decrease considerably faster, still the ones on this central line are larger than in other directions. With the selected square truncation we neglect these large coefficients far from the origin. Thus, from the truncated series we can expect a considerable deviation compared to the original function. The Fourier coefficients' imaginary parts of component $\tilde{\Pi}^{33}$ are plotted in the same way in panel (b). Here, the functions s_x and s_y are multiplied in real-space and their sharp features "add up". Consequently, the decay is even slower in both directions. A truncation with $|m_\rho| \leq 20$ neglects (too) much information, especially for Lalanne's PML formulation.

The imaginary parts of the corresponding reconstructed components are plotted throughout the unit cell below the original real-space distributions in Fig. 8.18 and Fig. 8.19 for $\tilde{\Pi}^{11}$ and $\tilde{\Pi}^{33}$, respectively. Here, the color scales are linear and saturated at values ± 3 in order to be able to see the details. It is clearly observable that the truncated Fourier series is unable to correctly represent the imaginary part of the coordinate transformation. The physical domain in the center of the plot where the function is supposed to be zero is dominated by oscillations. The behavior is similar to the oscillations we have seen before for the reconstructed functions f_ρ, but this time the oscillation amplitudes are

[5] Please keep in mind that we normalized all dimension to the lattice constant a

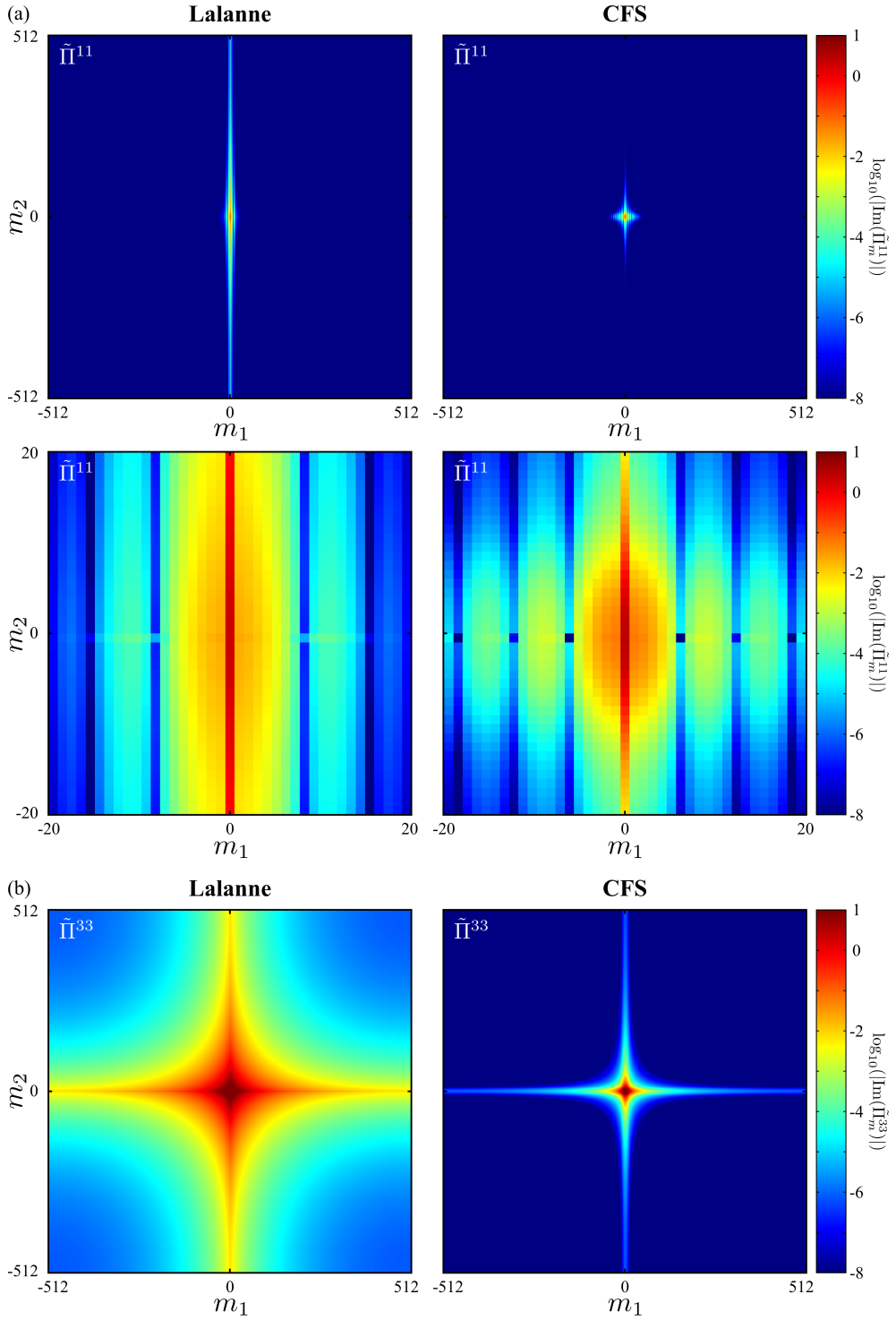


Figure 8.17.: *K-space representation of the PML transformation operator $\underline{\mathbf{\Pi}}$'s imaginary parts. The pseudo-color plots depict the magnitude of the Fourier coefficients on a logarithmic scale obtained by a two-dimensional FFT with $N_{\mathrm{fft}} = 1024$. The k-space origin is in the center of the plots. Panel (a) shows component $\tilde{\Pi}^{11}$ with all calculated coefficients in the top row, and the truncated sets with $|m_\rho| \leq 20$ in the bottom row. Panel (b) depicts the coefficients of component $\tilde{\Pi}^{33}$.*

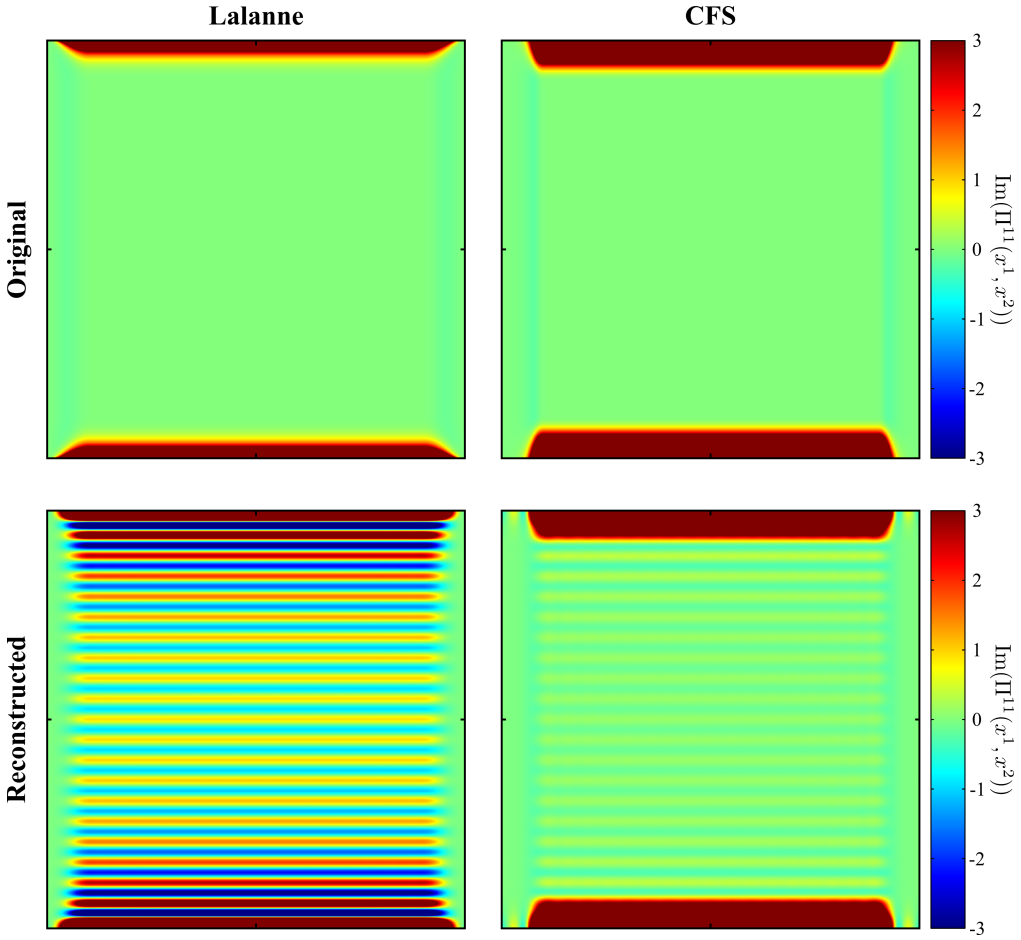


Figure 8.18.: *Real-space representation of the imaginary part of* Π^{11} *within the unit cell. Top row: original function, bottom row: reconstructed function from* $M = 1681$ *Fourier coefficients (*$|m_\rho| \leq 20$*, square truncation). The color scale is cut to highlight the details.*

much larger. In component $\tilde{\Pi}^{33}$ the oscillations in the two directions amplify each other mutually. Hence, the effective permittivity which the electromagnetic field encounters is a pattern of alternating lossy and gain material. Depending on the concrete mode, the geometry, the wavelength and several other parameters, the mode overlap with the structure will either be dominated by gain or loss in a highly unpredictable way. This is exactly what we observed for the imaginary parts of the effective refractive indices.

The oscillations for the PML parameters we choose in the convergence calculations above in Sec. 8.3.4 and Sec. 8.3.5, which is in particular $d = 1.001$, show qualitatively the same behavior, but the oscillations are even increased by a factor of 10^3 compared to what is plotted in Fig. 8.18 and Fig. 8.19. This is due to the even steeper tangent-like transformation function of the Lalanne PMLs near $\pm d/2$.

These k-space features with the prominent contributions from the axes, which holds similarly true for the real parts (not shown), might also explain the jumps in the convergence plot for certain numbers of modes M we mentioned before. We argued that the jumps occur exactly when another truncation

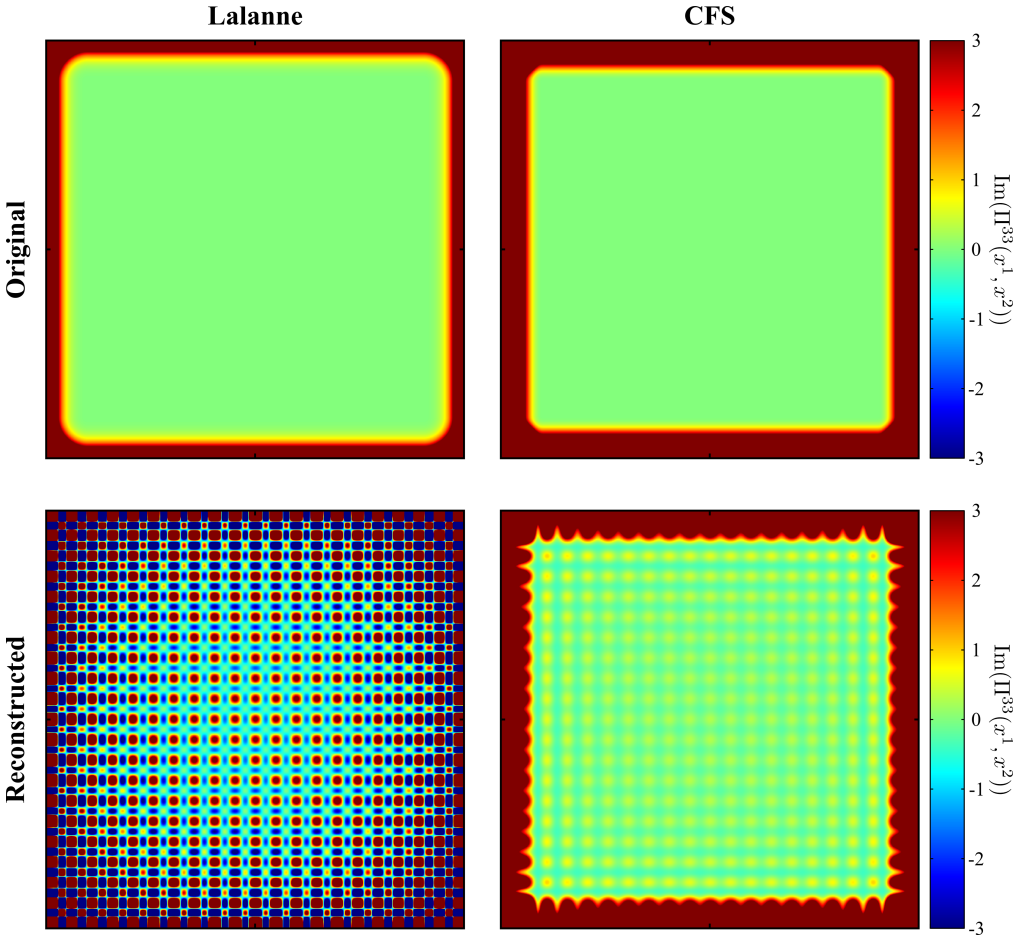


Figure 8.19.: *Real-space representation of the imaginary part of* Π^{33} *within the unit cell. Top row: original function, bottom row: reconstructed function from* $M = 1681$ *Fourier coefficients (*$|m_\rho| \leq 20$*, square truncation). The color scale is cut to highlight the details.*

order on the k-space axes joins the set of retained orders. Since the corresponding coefficients are very large, quite an amount of energy must be reshuffled from other plane waves into the new orders. Argued on the real-space level: With every retained Fourier component on the k-space axes, the oscillation pattern will encounter a considerable change and the modes "see" quite a different structure. This apparently results in a abrupt change of the modes' propagation constants.

In conclusion, we find that the PMLs in the literature, Lalanne and polynomially graded CFS, are both not well-suited for the covariant formulation of FMM (equation-transform real-space strategy). Because of the more appropriate curvature of the polynomials, the CFS-PML seem to perform slightly better than Lalanne's PMLs. The crucial difference between structure-transform and equation-transform strategies is that the former only Fourier transforms smooth bounded functions f_ρ, whereas the latter transforms their inverse s_ρ. As we have seen, an optimal representation is neither guaranteed by one nor the other. With these considerations, we could argue why the convergence plots exhibit the large jumps. Last not least, we could give strong evidence that the sign flip

problem of the effective indices' imaginary parts stems from the inaccurate numerical representation of the coordinate transformation.

The results imply two improvements for the future: On the one hand, we think that specifically tailored functions s_ρ which are easily represented by truncated Fourier series could help to overcome the issues. Instead of a tangent-like or polynomial grading which both result in sharp features at the unit cell boundaries, a Gaussian-like curvature seems advisable. The second improvement regards the truncation scheme: The coefficients on the k-space axes tend to decrease much slower than the others and, thus, should be retained with a higher priority. This implicates a star- or diamond-like truncation scheme where the truncation order on the axes is much larger than in other directions. Meanwhile, we show in Chap. 9 that, despite the described challenges, the method is well suited for the accurate and efficient simulation of large fiber-based devices.

Chapter 9.
9 Applications

This final chapter is concerned with the application of the developed numerical simulation tools to interesting real problems. The first part in this chapter, Sec. 9.1, presents the results of a rigorous theoretical analysis of an experimentally realized, periodic nano-photonic system, namely a woodpile structure with a complete bandgap in the region of visible light. This could be treated using the standard FMM implementation. The second part of this chapter, Sec. 9.2, then turns to the theoretical design of an aperiodic fiber-based long period grating mode coupler. Here, we demonstrate the entire toolkit of available FMM extensions developed in this work.

9.1. Woodpile Photonic Crystal with a Complete Bandgap in the Visible

In this section we present the results of a theoretical investigation of a woodpile photonic crystal structure experimentally fabricated by Froelich *et al.* [9]. The project aimed at a complete photonic bandgap in the visible range of the electromagnetic spectrum. To accomplish this challenging task, a resolution enhanced direct laser writing (DLW) laser lithography technique was used which is inspired by stimulated emission depletion (STED) as used in microscopy [8]. With this technique the effective dimension of the laser focus can be reduced to values way below the usual resolution limit of the used laser light ($\lambda \approx 800\,\mathrm{nm}$). With the DLW setup a polymeric template is written which is subsequently double inverted into titania ($\mathrm{TiO_2}$) with the help of an intermediate sacrifical zinc-oxide (ZnO) negative. ZnO and $\mathrm{TiO_2}$ are deposited in an atomic layer deposition (ALD) process from the gas phase. Scanning electron microscope pictures of the produced titania woodpile structure are shown in Fig. 9.1(b) and (c). Due to the conformality of the ALD process, the rods exhibit unavoidable small air voids in the center.

9.1.1. Setup

The structure is schematically depicted in Fig. 9.1(a) including the definition of the geometric parameters. The total structure has an outer dimension of $135\,\mu\mathrm{m} \times 70\,\mu\mathrm{m}$ and totals 33 layers, which

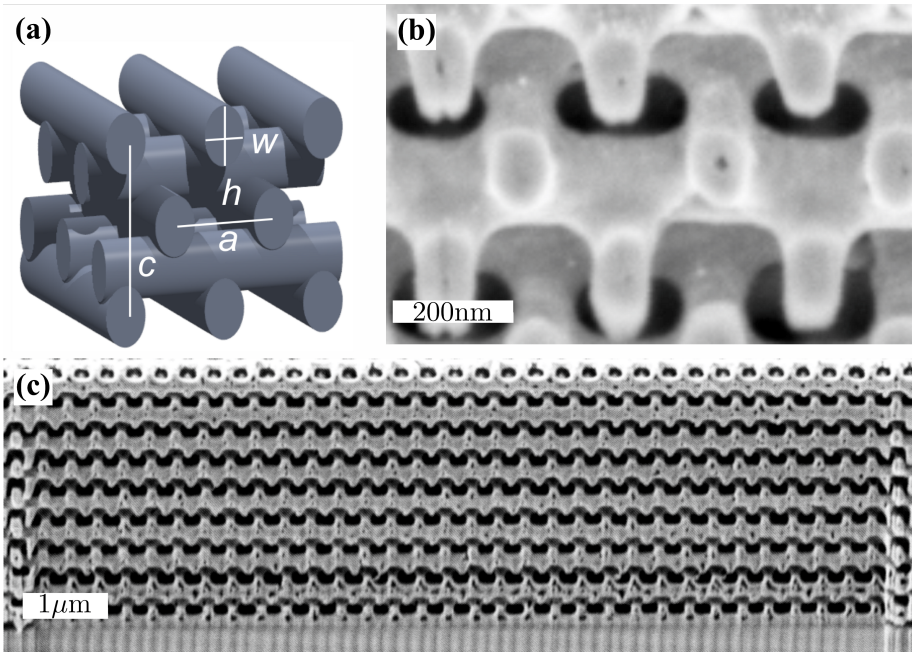


Figure 9.1.: *Three-dimensional titania woodpile. (a) Schematic picture of the geometry with a definition of the structure parameters. (b) Close-up electron micrograph of the titania woodpile after focused-ion-beam (FIB) milling to reveal its interior. The perfectly aligned rods are visible. (c) Overview of the FIB cut. The viewing angle is* $55.5°$. *Picture adapted from [9].*

corresponds to eight unit cells, containing 4 layers each, and a top layer.[1] It sits on top of a glass substrate ($\varepsilon = 2.3104$). The main propagation direction in the simulation, the x^3-axis of our coordinate system, is vertically from top to bottom. The x^1-axis is parallel to the axes of the rods in the top layer. The square in-plane (transversal) unit cell of lattice constant $a = 310\,\mathrm{nm}$ features one rod per layer with an equivalent in-plane spacing a. The height of one unit cell is given by compression length c. The vertical long axis of the elliptic rods is given by height h and their horizontal short axis by width w. The latter three parameters are determined by the analysis of the SEM image to be in the order of $c \approx 419\,\mathrm{nm}$, $h \approx 168\,\mathrm{nm}$, and $w \approx 99\,\mathrm{nm}$. Ultimately, they remain to be determined as exactly as possible from the comparison between experimental and theoretical transmittance results. This is reasonable because the FIB cut only represents a local measurement, and the sample clearly features small variations on the nanometer scale. The real space lattice of the structure is a compressed fcc lattice.

Since the air voids in the center of the rods can neither be measured reliably nor rigorously simulated, we estimate their volume fraction to be roughly 16 percent and arithmetically average the titania's permittivity values accordingly. Concerning the titania material parameters, the real and imaginary

[1] For historic reasons the simulational unit cell cuts the rods at odd positions, therefore the top layer is not one additional rod but only 0.675 rods. Hence, the theoretical setup is in total 32.675 layers. However, due to fabrication tolerances and for a better connection of the structure to the substrate, the experimental structure is also cut a little bit. With this many layers, the expected difference is a small shift of the Fabry-Perot resonances due to a different effective tickness of the structure.

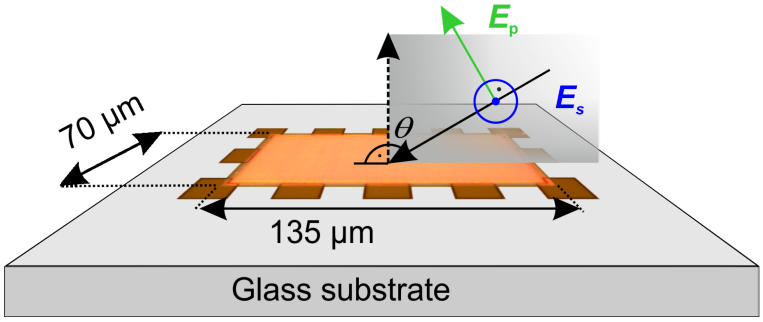


Figure 9.2.: *Blend of a true-color optical micrograph of the titania woodpile with the geometry for the optical transmission measurements. The angle of incidence with respect to the surface normal (dashed line) as well as s- and p-polarization of the incident light are indicated. Picture adapted from [9].*

part of the refractive indices were measured on an equivalent thin film sample. However, to account for surface roughness of the rods, we modify the imaginary part of the titania. It was shown in Ref. [114] that scattering from fabrication imperfections does not follow the wavelength-dependence expected from pure Rayleigh-scattering but instead follows a λ^{-2}-scaling. Therefore, we choose the imaginary part of the refractive index as $\mathrm{Im}(n(\lambda)) = \mathrm{i} \cdot 7500 \cdot \lambda^{-2} \cdot \mathrm{nm}^2$.

Figure 9.2 depicts a blending of a true-color optical micrograph image of the sample with a sketch of the irradiation configuration. The plane of incidence is perpendicular to the rod axes of the top layer.

9.1.2. Simulation and Comparison to Measured Data

The simulation of the structure is performed with $M = 197$ plane waves for angles of incidence from $\theta = 0°$ to $\theta = 60°$ in steps of $5°$ equivalent to the measured data. The unit cell is sliced into 50 layers of equal thickness. The wavelength range under consideration reaches from $\lambda = 500\,\mathrm{nm}$ to $\lambda = 850\,\mathrm{nm}$ in steps of $1\,\mathrm{nm}$. In total we performed more than 100 different runs of the program in different stages of the project. In the end, the best agreement between measurement and simulation is found for parameters $c = 416\,\mathrm{nm}$, $h = 174\,\mathrm{nm}$, and $w = 103\,\mathrm{nm}$. The corresponding transmittance spectra are plotted below the measurements in pseudo-color plots on a logarithmic scale in Fig. 9.3 for s- and p-polarization. The agreement in both cases is quite good which can be attributed to the high quality of the sample. The region with transmittance lower than a threshold of one percent for all angles of incidence and polarizations is highlighted with the horizontal dashed lines. This threshold is a reasonable choice for the definition of a stopband, since it can be seen as a sufficient factor of suppression of photonic states for a finite photonic crystal structure.

A notable difference between theory and experiment could be observed in all tested parameter sets. The wavelength minimum of the dielectric band in p-polarization is characteristically shifted between the two datasets. In the experimental data it is observed at $\theta = 30°$ whereas in all theoretical calculations it lies at $\theta = 35°$. Nevertheless, the overall good agreement between experiment and FMM simulation indicates a high quality sample with the indicated geometric dimensions. The remaining step is the confirmation of the complete bandgap with MPB.

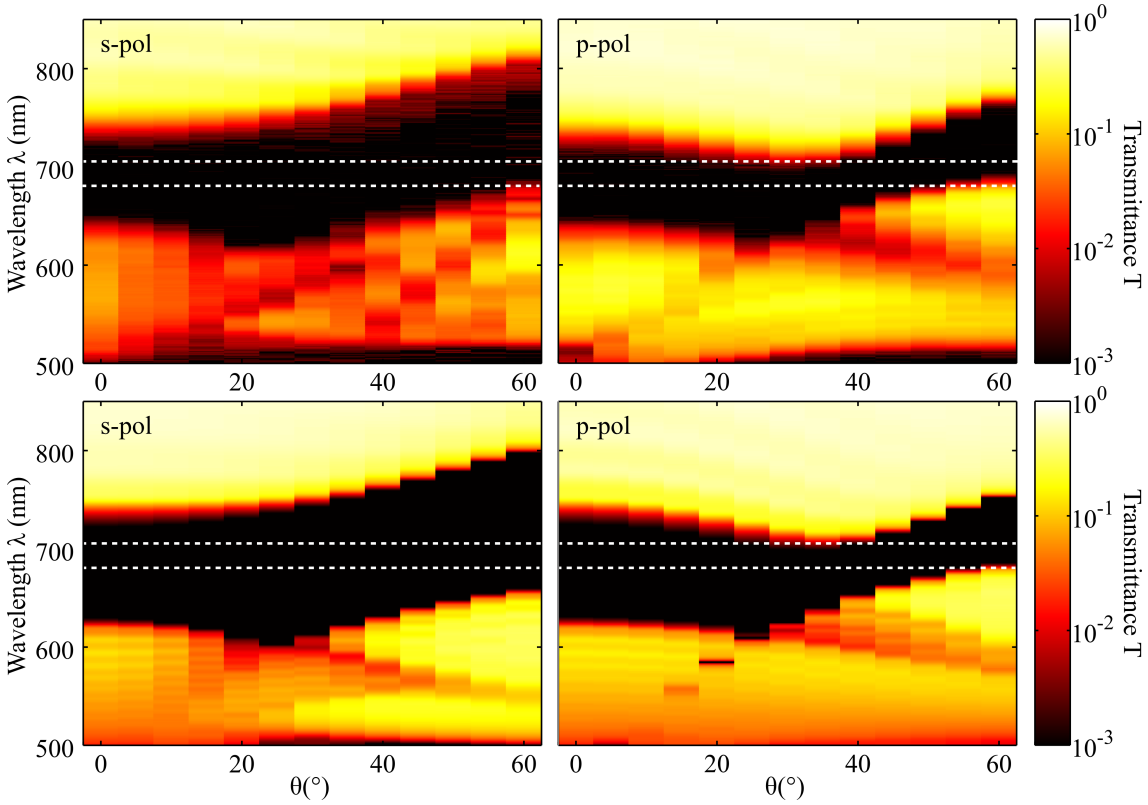


Figure 9.3.: *Comparison of measured (top) and simulated (bottom) transmittance spectra for s- and p-polarization and all angles of incidence on a logarithmic pseudo-color scale. The horizontal dashed lines indicate the region in which the transmission is lower than one percent for all angles and polarisations. The corresponding geometric parameters are given by* $a = 310\,\mathrm{nm}$, $w = 103\,\mathrm{nm}$, $h = 174\,\mathrm{nm}$, *and* $c = 416\,\mathrm{nm}$.

9.1.3. MPB Simulations

For the bandstructure simulations we performed the calculation with a discretization resolution of 256 sampling points and a mesh size of 20 using MPB. The high symmetry points for the stretched bcc lattice have been calculated analytically. They are given in Tab. 9.1. The refractive index for the titania rods was taken at a wavelength of $\lambda = 701\,\mathrm{nm}$ with a value of $n \approx 2.31$. For the geometry parameters from above, these calculations confirmed a complete photonic bandgap ranging from $694.2\,\mathrm{nm}$ to $706.5\,\mathrm{nm}$ with a gap to midgap ratio of 1.75 percent. This gap is completely in the visible part of the electromagnetic spectrum. The corresponding banddiagram is plotted in Fig. 9.4. The yellow shaded area indicates the measured wavelength range and the blue shaded region indicates the wavelength range where the measured data is below one percent. Finally, the red region highlights the complete photonic bandgap. The banddiagram only shows the most interesting paths on the edge of the irreducible Brillouin zone. In the simulation we considered many more. The limiting point from above is at the L-point, and from below near the W' point on the L-W'-line. With this calculation we have completed the theoretic calculations for the titania woodpile and

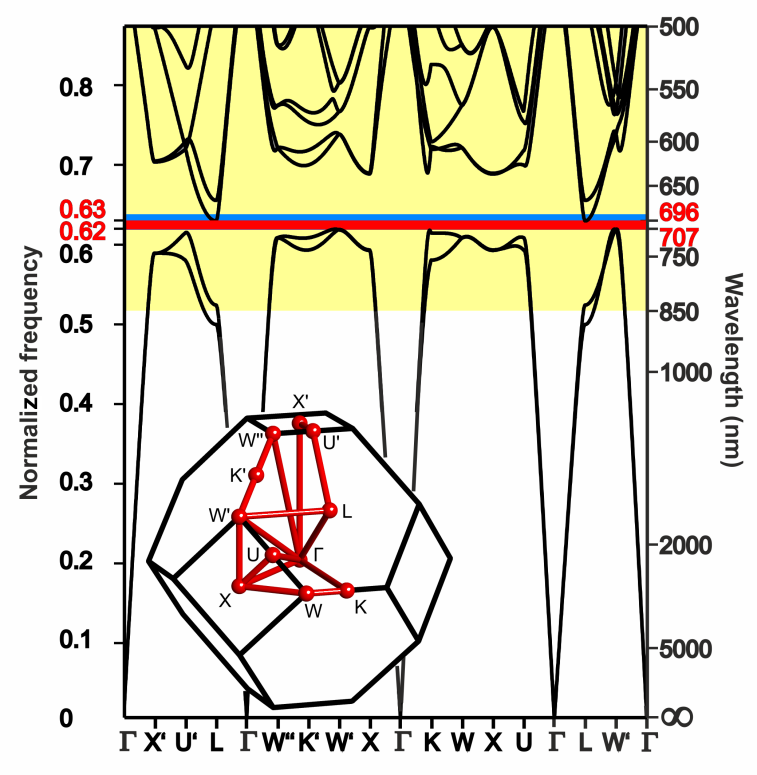


Figure 9.4.: *Complete three-dimensional photonic band gap in the visible. Depicted is the photonic banddiagram computed for the geometry parameters confirmed by the FMM. The complete three-dimensional photonic band gap between* $694.2\,\mathrm{nm}$ *and* $706.5\,\mathrm{nm}$ *is highlighted by the red area, the frequency region for which experimental and theoretical data is available is highlighted in yellow, and the region for which the measured transmission is below the threshold of one percent for all angles and polarisations is shown in blue. The inset shows the Brillouin zone: Red lines indicate paths covered in the banddiagram. Picture adapted from [9].*

could clearly provide strong evidence for the experimental realization of the first complete three-dimensional photonic bandgap in the visible.

Point	Real-space coordinate	K-space coordinate
L	$\pi\left(1,1,\frac{1}{s}\right)$	$0.5\left(1,1,1\right)$
X	$2\pi\left(1,0,0\right)$	$0.5\left(0,1,1\right)$
X'	$\frac{2\pi}{s}\left(0,0,1\right)$	$0.5\left(1,1,0\right)$
K	$2\pi\left(s+\frac{1}{2s},s+\frac{1}{2s},0\right)$	$0.25\left(1+\frac{1}{2s^2}\right)\left(1,1,2\right)$
W	$\frac{2\pi}{s}\left(s,\frac{1}{2s},0\right)$	$0.5\left(\frac{1}{2s^2},1,1+\frac{1}{2s^2}\right)$
U	$\frac{2\pi}{s}\left(s,\frac{1}{4s},0.25\right)$	$0.125\left(1+\frac{1}{s^2},5,4+\frac{1}{s^2}\right)$
W'	$\frac{2\pi}{s}\left(s,0,0.5\right)$	$0.25\left(1,3,2\right)$
K'	$\frac{2\pi}{s}\left(s-\frac{1}{4s},0,0.75\right)$	$0.125\left(3,7-\frac{1}{s^2},4-\frac{1}{s^2}\right)$
W''	$\frac{2\pi}{s}\left(s-\frac{1}{2s},0,1\right)$	$0.25\left(2,4-\frac{1}{s^2},2-\frac{1}{s^2}\right)$
U'	$\frac{\pi}{s}\left(s-\frac{1}{2s},s-\frac{1}{2s},2\right)$	$0.125\left(6-\frac{1}{s^2},6-\frac{1}{s^2},4-\frac{2}{s^2}\right)$

Table 9.1.: *List of analytically derived distinct points of the compressed woodpile's first Brillouin zone. The parameter $s = c/(\sqrt{2}a)$ denotes the compression factor of the fcc real-space lattice, and $V_E = s/4$ is the volume of the primitive cell. The real-space basis vectors are given by $\mathbf{a}_1 = 0.5(0,1,s)$, $\mathbf{a}_2 = 0.5(1,0,s)$, and $\mathbf{a}_3 = 0.5(1,1,0)$. The reciprocal lattice vectors are given by $\mathbf{a}^1 = 2\pi(-1,1,1/s)$, $\mathbf{a}^2 = 2\pi(1,-1,1/s)$, and $\mathbf{a}^3 = 2\pi(1,1,-1/s)$.*

9.2. Long Period Fiber Grating Mode Coupler

The numerical machinery which was developed, validated, and optimized within Chap. 6, Chap. 7 and Chap. 8 is now applied to the challenging problem of a fiber-based long period grating (LPG) mode coupler at telecom wavelengths. The system involves optimized adapted meshes including ASR, PML unit cell isolation, spatially inhomogeneous, fully anisotropic, dispersive material tensors, and symmetry reduction. The system's dimensions are a cross sectional square unit cell of size $a = 21\,\mu\mathrm{m}$ and an axial length of more than $8600\,\mu\mathrm{m}$ with over 800 layers.

9.2.1. Setup

The setup is schematically depicted in Fig. 9.5. The LPG we simulate is a (forward) grating designed to couple the fundamental waveguide mode incident from one side to a higher order guided mode leaving the device on the other side. It consists of a circular hollow core step-index fiber with a silica cladding ($r_{clad} = 7.5\,\mu\mathrm{m}$), where the core ($r = 2.15\,\mu\mathrm{m}$) is filled with a biaxially anisotropic liquid crystal (LC) material (Merck MDA-00-3969). The spatial orientation of the LCs can be manipulated by a transversal external electric field. From top and bottom, the fiber is enclosed by a periodic arrangement of electrodes with P periods of length Λ, the top electrodes covering half of the periods each, the bottom electrode the whole length of the grating. When the electrodes are biased with an external voltage V_{ext} between top and bottom, the core of the fiber is exposed to a x^3–periodic electric field. In general, this field is highly inhomogeneous and has to be derived in a self-consistent way, because of the interaction between the external field and the LC molecules.

The material properties of the LC material are characterized by the director vector field. The director

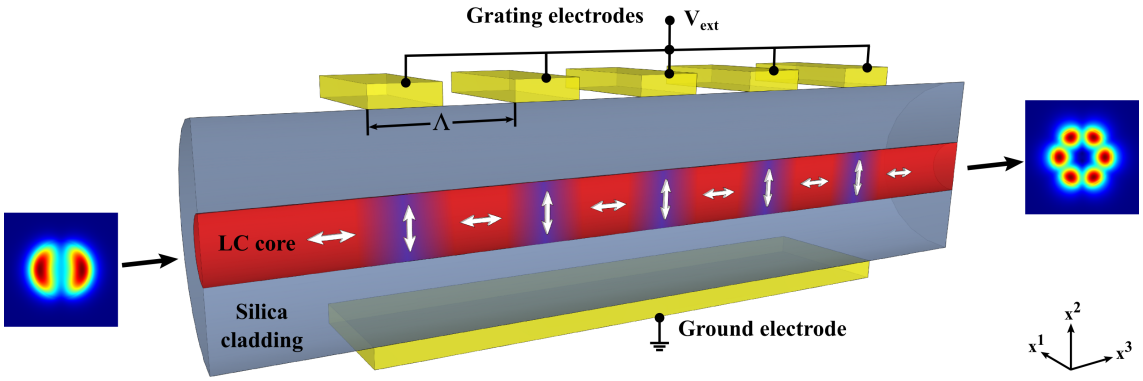


Figure 9.5.: *Schematic picture of the long period grating cut along the fiber axis through the x^2–x^3–plane. The core consists of a liquid crystal material, the cladding is made of silica. An external voltage V_{ext} applied between grating and ground electrode generates a periodic external electric field (field points along the x^2–direction) in the core with period Λ. The LC molecules tend to align with the external field which is sketched by the white double arrows, leading to periodic core material parameters indicated by the core color. By appropriate choice of Λ, the grating can be designed to couple an incoming HE_{11} mode (left) to an outgoing HE_{31} mode (right) at a desired resonance wavelength λ_{res}.*

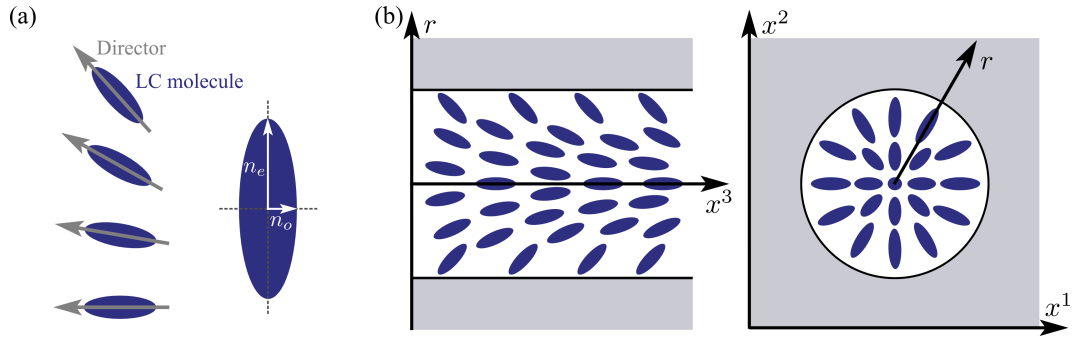


Figure 9.6.: *Schematic picture of liquid crystal molecules and their orientation within a hollow core fiber. (a) Cigar shaped LC molecules, director vector field, and definition of refractive indices. (b) Splay configuration. See text for details.*

points along the long axis of the cigar shaped LC molecule. The refractive index along this axis is called *extraordinary refractive index* n_e and varies approximately between 1.48489 at $\lambda = 1\,\mu\mathrm{m}$ and 1.48177 at $\lambda = 1.48\,\mu\mathrm{m}$. The refractive indices along the short axes of the LC molecules are given by the *ordinary refractive index* n_o which varies between about 1.68117 at $\lambda = 1\,\mu\mathrm{m}$ and 1.67359 at $\lambda = 1.48\,\mu\mathrm{m}$ wavelength. Infiltrated into a circular hollow core fiber, the LCs orient themselves with respect to the surrounding into a so-called *splay configuration* [115, 116], as schematically depicted in Fig. 9.6. The directors orient themselves parallel to the the r–x^3–plane. They form an $45°$ angle with the cladding interface, a $0°$ angle with the fiber axis, and vary smoothly in between. The splay configuration is rotationally symmetric. Application of an external voltage changes the director field out of this equilibrium position. With increasing external field strength the directors tend to align with the external field and slowly rotate towards this preferred direction.

Concerning the behavior of the molecule's director fields, to the best of our knowledge, there is no rigorous simulation available for their full three-dimensional calculation. However, we have access to two-dimensional self-consistent director field solutions obtained from a Finite Element simulation for varying external voltages [115, 116].

The external field reduces the rotational symmetry C_∞ (in our numerical example already reduced to $C_{4\mathrm{v}}$ because of the used quadratic lattice as mentioned in Sec. 8.3.2) of the unperturbed fiber to $C_{2\mathrm{v}}$. We will use this and reduce the simulated degrees of freedom accordingly (cf. Sec. 6.8.2). The whole size of the eigenproblem then shrinks by a factor of 4, the memory requirements by a factor of $4^2 = 16$, and the computation time for the eigenproblem by a factor of $4^3 = 64$. But then there are the four distinct symmetry subgroups which we simply call 1, 2, 3, 4 (cf. Sec. 6.8.2). Their solutions in general have to be determined by consecutive, independent calculations but considered all together. This leads to a theoretical speed up to a factor of $1/4 \cdot 4^3 = 16$. In practice, there is an overhead in the calculation for the preparation of the eigenproblem, but for $M = 997$ typical speed up factors of about 10 can be reached easily. Of course, it is not always necessary to consider all of the symmetry subgroups. In this section, the pictures, plots and effective refractive indices stem from the 4–subgroup unless otherwise noted.

Before we can begin with the grating design in Sec. 9.2.4, we have to know the propagation constants of the fiber's guided eigenmodes we want to couple for different external voltages. This is the task of the next section.

9.2.2. Guided Eigenmodes

The focus of this section is on the guided eigenmodes of the infiltrated fiber for different external voltages. To this end we solve the eigenproblem of a single layer.

A comparison between the electric field distributions obtained with the optimized adapted mesh and those obtained by standard FMM are depicted in Fig. 9.7. For this and all subsequent calculations we use the adaptive mesh optimized in Sec. 8.3.5 for the finite cladding system including PMLs with lattice constant $a = 21\,\mu\mathrm{m}$, $M = 997$ plane waves, and $N_{\mathrm{fft}} = 1024$ sampling points. The plot shows the HE_{11} and HE_{31} modes in the core region for an external voltage of $8V$ applied over the fiber diameter. Note that the adapted mesh improves first and foremost the representation of the field discontinuities at the core/cladding interface (cf. the insets). The barely visible asymmetry is owed to the used quadratic lattice as discussed in Sec. 8.3.2.

Next, the focus is put on the propagation constants. Figure 9.8 depicts the dispersion for the (a) unbiased and (b) homogeneously biased fibers in the wavelength range of interest. The main difference between both examined cases is the fact that the external field lifts some of the degeneracies. Figuratively speaking, degenerate modes with "more" electric field parallelly aligned to the director and, therefore, to the higher extraordinary refractive index of the LC move towards higher effective refractive index values. The obtained values for two of the guided modes which are marked with red circles (the two highest modes of symmetry 4, black dashed lines) are tabulated in Tab. 9.2 for different external voltages. We will need them below for the design of the grating period. With increasing voltage the reorientation of the LC molecules is small at first, but then increases more than linearly.

It should be noted that the material changes for different applied external voltages are rather subtle and so are the variations in the field distribution and propagation constants. In a visual comparison between the $0V$ and $8V$ cases, the differences would not be visible. Nevertheless, they suffice to form a diffraction grating with predefined properties.

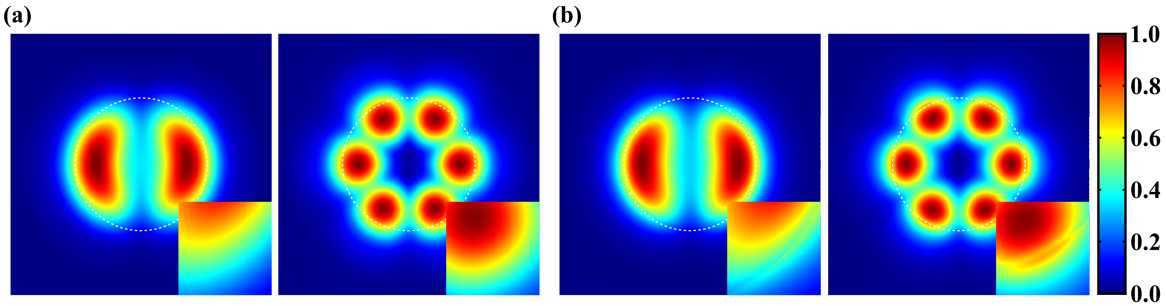


Figure 9.7.: *Normalized electric field distribution ($|\mathbf{E}|$) in the core region for the HE_{11} and HE_{31} guided eigenmodes of the externally biased LC filled anisotropic fiber ($8V$) plotted in the x^1–x^2–plane. Panel (a) shows the fields obtained with standard FMM and panel (b) the fields simulated with the adapted optimized mesh. All field strengths are normalized to one. The insets focus on the core/cladding interface where the transversal field components should be discontinuous.*

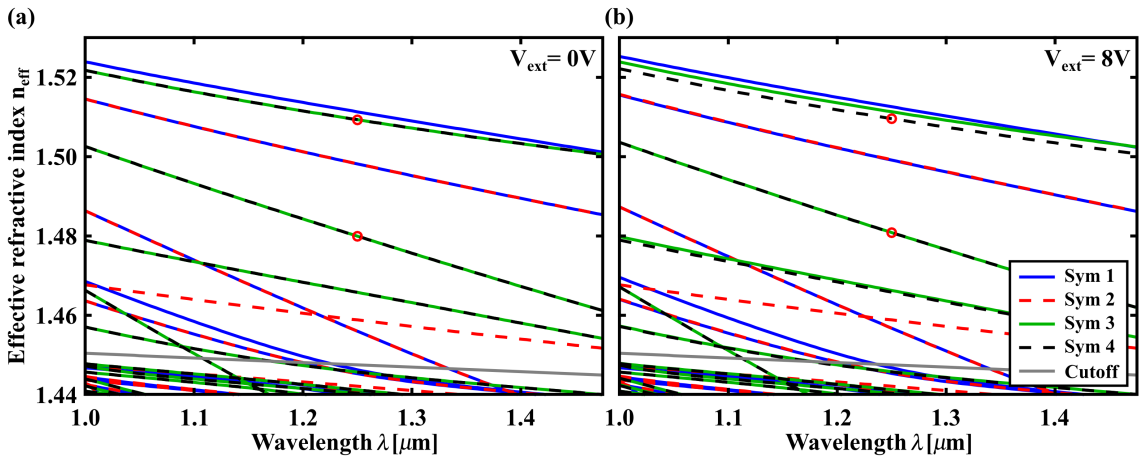


Figure 9.8.: *Dispersion spectra for the LC filled fibers color separated by their symmetry subgroups. (a) Dispersion spectrum for* $0V$*, (b) for* $8V$ *external voltage. The values marked by red circles are presented in Tab.* 9.2.

9.2.3. Structure Decomposition

For the simulation of the whole LPG with the FMM we make use of the structure's periodicity and decompose it along the fiber axis into P equivalent z-cells, each containing one full period. The z-cell is in turn decomposed into L layers of thickness t_l, $l = 1, \ldots, L$, where the permittivity in each layer is assumed to be homogeneous in the propagation direction x^3. As a consequence of the decomposition into a stack of slices, we have to choose the LC material model in accordance with our layers. This means we homogenize the director fields along the x^3 direction in each layer as well and calculate them as a stack of two-dimensional systems with varying external voltage. A detailed description of how the director fields are calculated is given in Refs. [115, 116].

We consider two approximations with different numbers of layers within the z-cell. The simplest z-cell consists of $L = 2$ layers, one being the slice of the fiber containing the electrode, the other being the slice of the fiber containing the gap between the electrodes. This first and rather crude approximation completely neglects the smooth variation in field strength, but is compatible with the

Mode	V_{ext}	n_{eff}	Mode	V_{ext}	n_{eff}
HE_{11}	0V	1.509312206	HE_{31}	0V	1.479930564
	2V	1.509316616		2V	1.479977244
	4V	1.509364198		4V	1.480145316
	6V	1.509447738		6V	1.480430684
	8V	1.509574112		8V	1.480840860

Table 9.2.: *Effective refractive indices for the liquid crystal infiltrated fiber calculated with* $N = 997$ *plane waves and the optimized adapted mesh at* $\lambda = 1.25\,\mu\text{m}$.

available material models. By setting the external voltage between top and bottom electrodes to 8V, we thus make the so-called *two layer approximation* — one layer containing the electrode with $V_{\mathrm{ext}} = 8V$ and an equally thick layer containing the gap with $V_{\mathrm{ext}} = 0V$. The input and output regions can be modeled by $L_{\mathrm{I}} = L_{\mathrm{O}} = 1$ additional unbiased layers of arbitrary thickness each. The solutions of the whole system can then be easily obtained by solving the $L + L_{\mathrm{I}} + L_{\mathrm{O}} = 4$ *different* quasi two-dimensional, independent layer subsystems for their eigenmodes, and recombine them afterwards considering the continuity conditions at the interfaces between adjacent layers by the scattering matrix algorithm (cf. Sec. 5.3).

The approximation from above not only totally neglects the smooth variation in the field profile, but does not respect the fact that the electric field decays with $1/r^2$ and, thus, will barely reach a zero external voltage domain in the small gap between two electrodes. An enhanced approximation of the periodic LC arrangement can be achieved by an electrostatic simulation of the external field strength in a homogeneous, isotropic core (like the step-index fiber in Sec. 8.3) in order to obtain a more realistic field profile. This enables us to introduce several layers with intermediate homogenized external field values. We are fully aware that this is still a rough approximation. However, it seems the best we can do with a two-dimensional LC material model, and the result surely serves as a proof of principle.

For the electrostatic simulation with COMSOL Multiphysics [117] we estimated a grating period of about $\Lambda = 43\,\mu\mathrm{m}$ and a spatial coverage of the upper electrode with regard to grating pitch of 50 percent. Using periodic boundary conditions in the direction of the fiber axis and an applied electrode voltage of $V_{\mathrm{ext}} = 8V$, we calculated the x^3–profile of the E_2 component in the center of the core. Since we are not interested in the field values but only in the spatial profile so as to determine the layer thicknesses t_l, we renormalize the obtained values such that the field strength maximum is $8V$ applied over the diameter of the fiber. The associated curve is plotted in Fig. 9.9(b). Similarly, we obtain profiles for the input and output regions on the left and right sides of the LPG, which is depicted in Fig. 9.9(a). From the displayed approximated curves we adopt the fraction of the grating period as thickness for the corresponding layers in our FMM simulation. The obtained values are denoted in the plots. Thus, the enhanced multi-layer approximation describes the grating period with $L = 4$ layers per grating period, $L_{\mathrm{I}} = 4$ input region layers, and $L_{\mathrm{O}} = 4 + 1$ output region layers, which makes a total of 13 layers and 13 eigenproblems to be solved. The additional layer in the output region is necessary, because the last electrode constitutes only half a grating period.

To stress this once more, the FMM is capable of approximating the system in x^3–direction as good as necessary by using an increasing number of layers. The simulation time scales only linearly with the number of distinguishable layers L. In this particular case our knowledge of the three-dimensional material properties is the limiting factor.

9.2.4. Designing the Grating Period

The essential step in the design of the LPG is the determination of the grating period Λ for a given resonance wavelength λ_{res}. In literature, e.g., in Ref. [118], one predominantly finds the relation

$$\lambda_{\mathrm{res}} = \left(n_{\mathrm{eff},1} - n_{\mathrm{eff},2}\right) \Lambda \,, \tag{9.1}$$

with $n_{\mathrm{eff},1}$ and $n_{\mathrm{eff},2}$ being the effective refractive indices of the launched and coupled-to mode, respectively. While this relation gives a good approximation to the resonance wavelength (or the

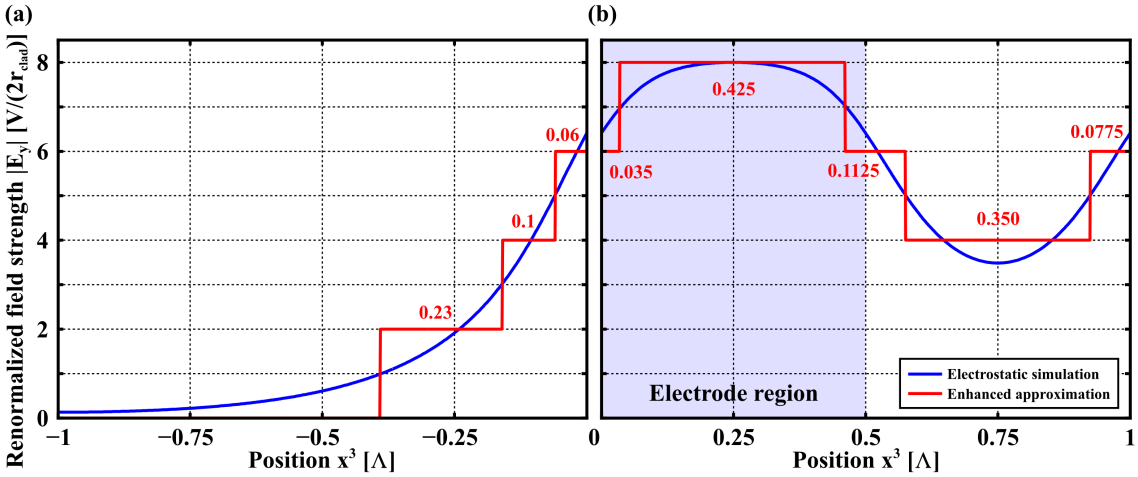


Figure 9.9.: *Electrostatically simulated $|E_y|$ field profile on the fiber axis for a LPG ($\Lambda = 43\,\mu\mathrm{m}$) with isotropic core as in Sec. 8.3 at wavelength $\lambda = 1.25\,\mu\mathrm{m}$. The rescaled profile and the corresponding enhanced approximation are depicted in panel (a) for the input/output region, and in panel (b) for one grating period. The top electrode covers the region from $x^3 = 0.0\Lambda$ to $x^3 = 0.5\Lambda$ (periodic boundary conditions in x^3) and is biased with $V_{\mathrm{ext}} = 8V$. The resulting layer thicknesses d_l are denoted next to the curve sections (red).*

grating period accordingly), it is not accurate because it describes a grating of point scatterers — infinitely small electrodes in our case, i.e., infinitely thin $8V$ layers in a two layer approximation. For fiber grating purposes we would like to have a more accurate relation which is able to handle the finite thickness of the electrode or, more precisely, the proper profile of the material properties as approximated by two or more constituent layers.

Hence, we start with a derivation from the basic physical condition: Resonance between two co-propagating (or counter-propagating) modes occurs if their phases match after a whole grating period

$$\phi_2(\Lambda) = \phi_1(\Lambda) + m \cdot 2\pi\,, \quad m = 0, \pm 1, \pm 2, \dots\,, \tag{9.2}$$

where ϕ_i are the phases of the modes, and m labels the diffraction orders. The phase of a mode is obtained by integrating its propagation constant over one period

$$\phi_i(\Lambda) = \int_0^{\Lambda} \beta_i(x^3)\mathrm{d}x^3\,, \tag{9.3}$$

which leads in the illustrative case of a decomposition into two finite constituent layers A and B with thicknesses t_A and t_B to

$$\Big(\beta_{1,A}t_A + \beta_{1,B}t_B\Big) - \Big(\beta_{2,A}t_A + \beta_{2,B}t_B\Big) + m \cdot 2\pi = 0\,. \tag{9.4}$$

Assuming co-propagating modes $\beta_1 > \beta_2 > 0$, a resulting negative diffraction order $m = -1$ (first order is preferred to higher orders because of larger coupling strength), and two equally thick layers

$t_A = t_B = \Lambda/2$, we get

$$\lambda_{\mathrm{res}} = \left(\frac{n_{\mathrm{eff},1,A} + n_{\mathrm{eff},1,B}}{2} - \frac{n_{\mathrm{eff},2,A} + n_{\mathrm{eff},2,B}}{2} \right) \Lambda \,, \tag{9.5}$$

where we used $\beta_i = 2\pi \cdot n_{\mathrm{eff},i}/\lambda_{\mathrm{res}}$. The result is neat, as it simply involves the weighted arithmetic averages of the propagation constants instead of the propagation constants of the unperturbed waveguide. The most general relation for an arbitrary number of layers is then given by

$$\lambda_{\mathrm{res}} = \left(\bar{n}_{\mathrm{eff},1} - \bar{n}_{\mathrm{eff},2} \right) \Lambda \,, \tag{9.6}$$

with the arithmetically averaged effective refractive index

$$\bar{n}_{\mathrm{eff},i} = \phi_i \cdot \frac{\lambda_{\mathrm{res}}}{2\pi\Lambda} \,, \tag{9.7}$$

and ϕ_i given by Eq. (9.3).

With the help of this result it is possible to calculate the grating periods for the two layer approximation. The resonance wavelength is designed to $\lambda_{\mathrm{res}} = 1.25\,\mu\mathrm{m}$ for a coupling between the arbitrarily picked modes HE_{11} and HE_{31}. Taking the values from Tab. 9.2, the grating period for the two layer system ($0V$ and $8V$, $d_{0V} = d_{8V} = \Lambda/2$) is $\Lambda_{\mathrm{2Layer}} = 43.01823212\,\mu\mathrm{m}$. Similarly, the enhanced multi-layer approximation grating period Λ_{Enh} can be calculated using the parameters displayed in Tab. 9.2 and the thicknesses denoted in Fig. 9.9 (b). One obtains the grating period $\Lambda_{\mathrm{Enh}} = 43.15243678\,\mu\mathrm{m}$ from Eq. (9.6).

9.2.5. Mode Conversion

The transmittances of both simulated systems are depicted in Fig. 9.10 and Fig. 9.11 color-coded over the wavelength λ and number of grating periods P. For the two layer approximation in Fig. 9.10 the HE_{11} mode completely couples to the HE_{31} mode at the designed resonance wavelength after 33 grating periods with a mode conversion factor of 99.13 percent. With an increasing number of grating periods, the mode couples back and forth with a beating length of 67 grating periods. The dips in the HE_{11} transmittance spectrum are higher order resonances with cladding modes. The mode conversion factor as well as the whole transmittance slowly decreases because of the guided modes' remaining imaginary part of the propagation constants due to the PML absorption. Investigations without PMLs in a periodic arrangement show the same behavior, but lack the overall absorption. Even without PMLs the coupling to guided modes in neighboring fibers is negligible. The transmittance into *every* single eigenmode of the system naturally comes out of the same calculation as well.

A simulation with the more accurate external field profile along the fiber axis of the enhanced multi-layer approximation leads to the transmittance spectra shown in Fig. 9.11. With the properly adjusted grating period the simulation results show very similar behavior as before. Naturally, the higher order cladding mode resonances shift relatively to the main coupling resonance due to the altered phase relation. In addition, the beating length of the coupling resonance increases to 93 periods with the first full coupling to the HE_{31} mode at about 46 periods.

If we choose a higher maximal external voltage, we obtain many more resonances with other waveguide modes whose fringes tend to overlap and the nice distinct resonance with the HE_{31} mode gets

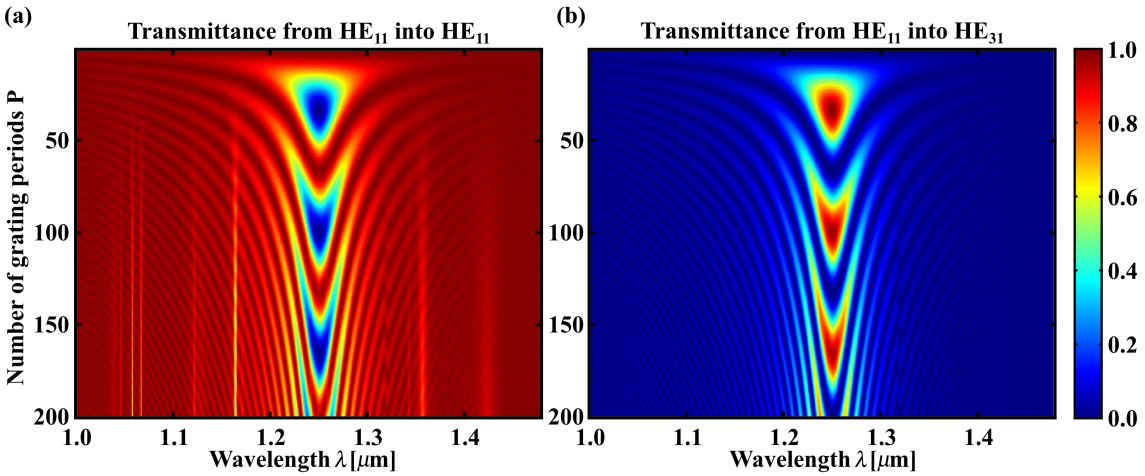


Figure 9.10.: *Transmittance spectra of the LPG in the two layer approximation for a varying number of grating periods. The grating is illuminated with mode* HE_{11} *and panel (a) shows transmittance into the same mode, whereas panel (b) depicts the coupling into mode* HE_{31} *(cf. Fig. 9.7). The coupling resonance is designed to a wavelength of* $1.25\,\mu\mathrm{m}$.

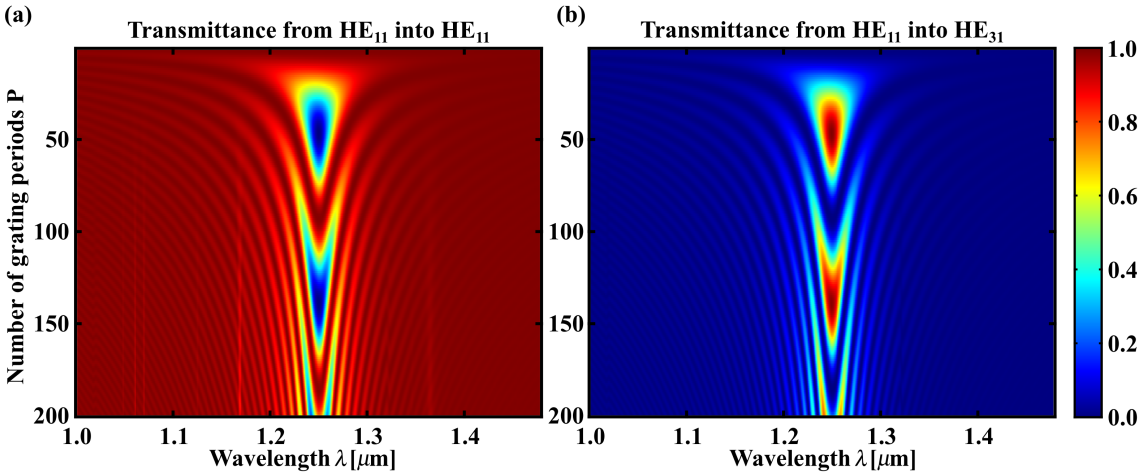


Figure 9.11.: *Transmittance spectra of the LPG in the enhanced approximation for a varying number of grating periods. The grating is illuminated with mode* HE_{11} *and panel (a) shows transmittance into the same mode, whereas panel (b) depicts the coupling into mode* HE_{31} *(cf. Fig. 9.7). The coupling resonance is designed to a wavelength of* $1.25\,\mu\mathrm{m}$.

polluted. If the external voltage is chosen smaller, the coupling gets weaker even though the grating effect in general is preserved. After all, it is not the amplitude of the perturbation which is crucial, but the grating period. Thus, $8V$ to $10V$ turn out to be a good compromise for this setup.

9.2.6. Computational Costs

In these calculations, we use acceleration by symmetry reduction of the problem, since the incident mode is a member of C2v symmetry group 4 (cf. Sec. 9.2.2). Studies of the two layer system show that, simulating the full symmetry problem (C4v), merely less than one percent of the light is coupled to the degenerate HE_{31} mode of C2v symmetry group 3. The remainder of the incident light is coupled to the HE_{31} mode of C2v symmetry group 4. This means that the coupling between symmetry groups, which occurs in layers with zero external voltage, is negligible. Therefore, it is possible to reduce the complexity of the whole problem and only consider group 4. Consequently, the number of elements in the eigenproblem matrix as well as the scattering matrix are reduced by a factor of 4^2. This massively speeds up the calculation time for each eigenvalue problem from about 700 seconds to 12.3 seconds (i.e., a factor of 56 vs. a theoretical factor of $4^3 = 64$) per wavelength and layer on a single core of an Intel Xeon X5660 processor.

Furthermore, the strengths of the scattering matrix approach are exploited: Separate scattering matrices for the input region (S_I), output region (S_O), and *one* z-cell (S_ZC) are prepared first. The latter contains the eigensolutions of the L layers of one grating period. The scattering matrix for a grating with P periods is then created by successively adding one further period ($S_{P\cdot\mathrm{ZC}} = S_\mathrm{ZC} \star S_{(P-1)\cdot\mathrm{ZC}}$). This is achieved by a comparably inexpensive S-matrix product (cf. Sec. 5.3.3). The total scattering matrix $S_\mathrm{T} = S_\mathrm{O} \star S_{P\cdot\mathrm{ZC}} \star S_\mathrm{I}$ is finally synthesized from the input, grating, and output scattering matrices and evaluated. Consequently, instead of the $(P \cdot L) + L_\mathrm{I} + L_\mathrm{O}$ layers of the whole structure, we have to calculate $L + L_\mathrm{I} + L_\mathrm{O}$ layer solutions only. The contributions from the remaining grating periods of the structure are obtained by recycling the scattering matrix of the z-cell.

However, there is even potential for a further reduction of computational expenses for this particular system which we did not exploit yet. For instance, layers with intermediate external voltage appear twice per period, input and output region are symmetric, and eigenmodes of a layer are independent of the layer's thickness. The latter means that layers with the same applied external voltage, but different thicknesses lead to the same eigenproblem. Taking everything into account, there are only five distinct layers and, thus, five eigenproblems that need to be solved, even for the LPG with 200 grating periods and a total of $P \cdot L + L_\mathrm{I} + L_\mathrm{O} = 200 \cdot 4 + 4 + 5 = 809$ computational layers.

The calculation of the whole spectrum is separated and parallelized into independent runs per wavelength using MPI [119]. A single run takes about 1638 seconds (27.3 minutes) for the enhanced approximation with 13 different layers and all 200 calculated different grating periods. Using symmetry reduction, most time is spent for the preparation of the eigenproblem with 33.7 seconds in each layer. After the preparation of the scattering matrix building blocks (35 seconds) the transmittance calculation for each additional grating period takes 4.9 seconds on average only.

10 Chapter 10. Conclusion and Outlook

In this thesis we presented the Fourier modal method and demonstrated its versatility and efficiency by the characterization, design and investigation of periodic and aperiodic photonic systems. Furthermore, we expedited several major projects.

First, we could demonstrate that the FMM is an essential numerical tool at the interface between experiment and bandstructure simulations for photonic crystals. It closes the fundamental gap in the evidence chain for complete photonic bandgaps. The angle- and polarization resolved transmittance and reflectance calculations can directly be related to the experimental spectroscopic measurements. Hence, the data analysis provides an immediate characterization of the fabricated structure. The method's efficiency and a parallel implementation allows for wide range parameter scans on large computer clusters. This enabled us to help with the selection of promising woodpile templates right from the start. Later on, the parameter studies narrowed down the uncertainty in the geometric features obtained from the scanning electron microscope images. With the good agreement between experiment and FMM simulations for a wide range of angles and both polarizations on the one hand, and the compliance between FMM and MPB on the other hand, we could provide strong evidence for the first successful engineering of a complete photonic bandgap in the visible.

Second, we demonstrated that the FMM is suitable and capable of rigorously calculating the guided eigenmodes of fiber systems and related large scale devices like the LPG mode coupler based on an liquid crystal infiltrated fiber. In order to reach this point, we worked on several extensions to the method.

On the foundation of curvilinear coordinates established by Sabine Essig, we extended the code to fully anisotropic material tensors and, as a consequence, large eigenproblems. In this context, new schemes for the determination of forward and backward propagating modes had to be established.

In collaboration with Jens Küchenmeister, we developed and explored the idea of smoothed and differentiable analytic adaptive coordinate meshes with the principles of transformation optics. We gave an intuitive and clear manual for analytic mesh constructions of all three types, which bases on the simple concept of linear transitions between characteristic specific lines. This scheme is easily extendable to more complex structures like crescents or trapezoidal slotted waveguides. In a rigorous study, we could show that the non-differentiable meshes known from literature give the most accurate results, provided the ASR parameters are correctly chosen. With the detailed analysis we gave, the large parameter scans for the mesh optimization will become obsolete in the future because suitable

mesh parameters can be analytically derived from the transformation functions.

The established covariant formulation of the FMM could be exploited for the implementation of streched-coordinate PMLs as well. While the FMM literature only knows Lalanne's type within the equation-transform k-space strategy, we integrated the complex coordinate stretching PMLs with the real coordinate transformations of AC and ASR in the structure-transform real-space strategy. Furthermore, we adopted CFS-PMLs from time domain methods. Our investigations suggest that the PML region should be chosen as large as possible but with respect to the structure to unit cell size ratio. A detailed convergence analysis was able to clarify that the PMLs work — albeit with certain restrictions.

None of the PML types seem optimally suited for truncated Fourier series, because strong mesh compressions create spurious modes in the guided modes region. This complicates the automated determination of guided modes which is usually essential for the launching. Furthermore, due to the introduced absorbing components, even the truly guided modes gain imaginary parts in the effective indices. Additionally, we observed the effect of random sign flips which lead to simulated transmittances above one. However, with a detailed analysis of the PML's contribution to the effective material parameters, we could give evidence for the origin of the effect. We concluded that character and curvature of the functions s_ρ of both PML types lead to oscillations of the imaginary parts arround zero within the physical domain. Depending on the concrete circumstances, the related gain effects apparently overcompensate the intrinsic losses within the PML region. Two possible resorts were proposed: Firstly, an improved truncation scheme in k-space which respects the dominant coefficients on the axes. And secondly, an optimized profile of the functions s_ρ which avoids oscillations of the imaginary part within the physical domain.

It is the first time that the combination of different coordinate transformations for perfectly matched layers and adaptive meshes with adaptive spatial resolution were successfully applied together with inhomogeneous, fully anisotropic, dispersive materials and the C_{2v} symmetry reduction in one simulation. The thoroughly optimized adapted mesh for the LPG mode coupler led to a reduction of the propagation constant's maximal relative error of almost three orders of magnitude. PML absorbing boundaries helped to inhibit cross-talk between neighboring unit cells.

With these improvements we accurately determined the propagation constants of two of the liquid crystal filled fiber's guided eigenmodes. These calculations enabled the design of the required grating period. In this context we improved the related formula for the resonance wavelength such that arbitrary grating profiles can be handled. The limiting factor in the simulation was the material model, because there is no three-dimensional self-consistent model available for liquid crystals. Instead, we resorted to two-dimensional material simulations. As a consequence, we approximated the material and grating profile along the fiber by two and five layers, whose effective external voltages were determined in an approximative electrostatic simulation. However, the number of layers per grating period is in principle not limited and can be increased up to the desired level of accuracy with only linear costs. The physical results showed a beating of the mode coupling strength with over 99 percent coupling efficiency. The scattering into the cladding or the surrounding was negligible. The outstanding strength of the method could be demonstrated by the efficient simulation of the system with up to 200 distinct grating periods with only little additional effort.

In conclusion, we could prove that both the FMM and the extended FMM in covariant formulation are well suited for simulations of periodic photonic crystals and for high aspect ratio simulations of aperiodic systems such as fiber devices, respectively.

Outlook

Throughout this work, we discovered many starting points for future projects. Here, we would like to give a brief overview of the most important ones.

- Implementation and testing of optimized PMLs: As mentioned above.
- Implementation and exploration of adapted k-space truncation schemes: As mentioned above.
- Reevaluation of eigenmode determination for realistic PCFs: On the basis of the gained experience with adapted meshes and fiber eigenmodes, in general, a new benchmarking against COMSOL would be worthwhile.
- Evaluation of Li's rules in two dimensions: There are indications that the inverse rule applied to type 1 problems converges equally badly as Laurent's rule. As a consequence, the complicated distinction between the two types might be obsolet. Instead, a use of the inverse rule in both directions should be evaluated.
- Application of Singular Fourier-Padé approximations for the field reconstruction: To this end, the SFP must be generalized to two dimensions. Especially, the incorporation of the jump locations must be solved consistently.
- Calculation of Maxwell stress-tensors: This is essential for the calculations of forces acting on particles. The basic obstacle to be overcome is the exact determination of the particle surface — especially for heavily staircased structures.
- Construction of nested meshes: The next step in the development of construction guidelines is the generalization of the established rules to topologically more difficult structures.
- Automated analytic mesh generation: The development of an automated mesher based on the established construction principles might be interesting for a larger audience. The mesher could also incorporate the automated choice of optimal ASR parameters. This could be particularly interesting in the context of realistic photonic crystal fibers.
- Automated adaptive mesh generation: An optimized adapted design of the gradient "potential" and an exploration of the shear term is desirable. This could reduce gratuitous mesh deformations.
- Three-dimensional coordinate transformations: The current code is by and large prepared for three-dimensional transformations. First meshes are constructed, but a rigorous testing remains an open point. The layer matching and staircasing needs to be explored. One of the first interesting problems could be the transformation of a sphere into a cuboid. As reference test one could calculate the scattering cross section (Mie theory).
- Redesign of the code: The introduction of an additional abstraction layer which coordinates and optimizes a dynamic scattering matrix construction from the XML configuration file is necessary. In this context an automatic decision for the best eigenproblem on a single layer basis could be implemented. Last not least, a graphical user interface could tremendously facilitate the usability.

Closing remark

We belive that we could provide a significant contribution to the progress of the Fourier modal method with this work — in particlar to coordinate transformations. Still, every answer to an open question induces several new questions and the story never ends. In this sense, we are curious to see where the future path will lead.

A Appendix A.

Fourier Transformation

A.1. Full Anisotropic Fields

Suppose we have the contravariant dielectric displacement D^ρ and magnetic flux density B^ρ. We are interested how these quantities transform under a Fourier transformation obeying Li's rules for Fourier transformations of products.

We start with the dielectric displacement in real space

$$D^\rho(x^1, x^2) = \epsilon^{\rho\sigma}(x^1, x^2) E_\sigma(x^1, x^2) \; . \tag{A.1}$$

The outwritten form is

$$D^1 = \varepsilon^{11} E_1 + \varepsilon^{12} E_2 + \varepsilon^{13} E_3 \tag{A.2a}$$
$$D^2 = \varepsilon^{21} E_1 + \varepsilon^{22} E_2 + \varepsilon^{23} E_3 \tag{A.2b}$$
$$D^3 = \varepsilon^{31} E_1 + \varepsilon^{32} E_2 + \varepsilon^{33} E_3 \; . \tag{A.2c}$$

Here, and in the following the spatial dependence is assumed but suppressed for brevity. We will perform Fourier transformations in dimensions x^1 and x^2 successively one after the other. Li's rules describe wich transformation rules should be chosen for the transformation of products of functions, namely the Laurent's rule for products of at most one discontinous function at a certain spatial point, and the inverse rule for functions with concurrent jump discontinuities.

Let us start with a Fourier transformation in x^1 direction. Thus, in Eqs. (A.2) we know for sure that D^1, E_2, and E_3 must be continuous across the whole unit cell. The task at hand is to reformulate Eqs. (A.2) is such a way that we only get products of functions which are of Laurant or inverse type. Starting with the first component we get

$$D^1 = \varepsilon^{11} \left[E_1 + \frac{\varepsilon^{12}}{\varepsilon^{11}} E_2 + \frac{\varepsilon^{13}}{\varepsilon^{11}} E_3 \right] \; . \tag{A.3}$$

Since D^1 is continuous, but ε^{11} is defnitely discontinuous, the term in the square brackets must also be discontinuous and its product with ε^{11} must be of inverse type. The products within the square brackets are of Laurent type. This implies

$$\begin{aligned}\tilde{\mathbf{D}}^1 &= \left[\!\left[\frac{1}{\varepsilon^{11}}\right]\!\right]^{-1}\tilde{\mathbf{E}}_1 + \left[\!\left[\frac{1}{\varepsilon^{11}}\right]\!\right]^{-1}\left[\!\left[\frac{\varepsilon^{12}}{\varepsilon^{11}}\right]\!\right]\tilde{\mathbf{E}}_2 + \left[\!\left[\frac{1}{\varepsilon^{11}}\right]\!\right]^{-1}\left[\!\left[\frac{\varepsilon^{13}}{\varepsilon^{11}}\right]\!\right]\tilde{\mathbf{E}}_3 \\ &=: \quad \underline{\mathbf{Q}}^{11}\,\tilde{\mathbf{E}}_1 \quad + \quad \underline{\mathbf{Q}}^{12}\,\tilde{\mathbf{E}}_2 \quad + \quad \underline{\mathbf{Q}}^{13}\,\tilde{\mathbf{E}}_3\,, \end{aligned} \tag{A.4}$$

where all field components and matrices[1] still depend on x^2.
Similarly, we obtain

$$\begin{aligned} D^2 &= \varepsilon^{21}E_1 + \varepsilon^{22}E_2 + \varepsilon^{23}E_3 \\ &= \left(\frac{\varepsilon^{21}}{\varepsilon^{11}}\right)\left[\varepsilon^{11}E_1 + \varepsilon^{12}E_2 + \varepsilon^{13}E_3 - \varepsilon^{12}E_2 - \varepsilon^{13}E_3\right] + \varepsilon^{22}E_2 + \varepsilon^{23}E_3 \\ &= \left(\frac{\varepsilon^{21}}{\varepsilon^{11}}\right)D^1 + \left(\varepsilon^{22} - \frac{\varepsilon^{21}\varepsilon^{12}}{\varepsilon^{11}}\right)E_2 + \left(\varepsilon^{23} - \frac{\varepsilon^{21}\varepsilon^{13}}{\varepsilon^{11}}\right)E_3\,, \end{aligned} \tag{A.5}$$

$$\begin{aligned} \tilde{\mathbf{D}}^2 &= \left[\!\left[\frac{\varepsilon^{21}}{\varepsilon^{11}}\right]\!\right]\left(\left[\!\left[\frac{1}{\varepsilon^{11}}\right]\!\right]^{-1}\tilde{\mathbf{E}}_1 + \left[\!\left[\frac{1}{\varepsilon^{11}}\right]\!\right]^{-1}\left[\!\left[\frac{\varepsilon^{12}}{\varepsilon^{11}}\right]\!\right]\tilde{\mathbf{E}}_2 + \left[\!\left[\frac{1}{\varepsilon^{11}}\right]\!\right]^{-1}\left[\!\left[\frac{\varepsilon^{13}}{\varepsilon^{11}}\right]\!\right]\tilde{\mathbf{E}}_3\right) \\ &\quad + \left[\!\left[\varepsilon^{22} - \frac{\varepsilon^{21}\varepsilon^{12}}{\varepsilon^{11}}\right]\!\right]\tilde{\mathbf{E}}_2 + \left[\!\left[\varepsilon^{23} - \frac{\varepsilon^{21}\varepsilon^{13}}{\varepsilon^{11}}\right]\!\right]\tilde{\mathbf{E}}_3 \\ &= \left[\!\left[\frac{\varepsilon^{21}}{\varepsilon^{11}}\right]\!\right]\left[\!\left[\frac{1}{\varepsilon^{11}}\right]\!\right]^{-1}\tilde{\mathbf{E}}_1 + \left(\left[\!\left[\varepsilon^{22} - \frac{\varepsilon^{21}\varepsilon^{12}}{\varepsilon^{11}}\right]\!\right] + \left[\!\left[\frac{\varepsilon^{21}}{\varepsilon^{11}}\right]\!\right]\left[\!\left[\frac{1}{\varepsilon^{11}}\right]\!\right]^{-1}\left[\!\left[\frac{\varepsilon^{12}}{\varepsilon^{11}}\right]\!\right]\right)\tilde{\mathbf{E}}_2 \\ &\quad + \left(\left[\!\left[\varepsilon^{23} - \frac{\varepsilon^{21}\varepsilon^{13}}{\varepsilon^{11}}\right]\!\right] + \left[\!\left[\frac{\varepsilon^{21}}{\varepsilon^{11}}\right]\!\right]\left[\!\left[\frac{1}{\varepsilon^{11}}\right]\!\right]^{-1}\left[\!\left[\frac{\varepsilon^{13}}{\varepsilon^{11}}\right]\!\right]\right)\tilde{\mathbf{E}}_3 \\ &=: \quad \underline{\mathbf{Q}}^{21}\,\tilde{\mathbf{E}}_1 \quad + \quad \underline{\mathbf{Q}}^{22}\,\tilde{\mathbf{E}}_2 \quad + \quad \underline{\mathbf{Q}}^{23}\,\tilde{\mathbf{E}}_3\,, \end{aligned} \tag{A.6}$$

$$\begin{aligned} D^3 &= \varepsilon^{31}E_1 + \varepsilon^{32}E_2 + \varepsilon^{33}E_3 \\ &= \left(\frac{\varepsilon^{31}}{\varepsilon^{11}}\right)\left[\varepsilon^{11}E_1 + \varepsilon^{12}E_2 + \varepsilon^{13}E_3 - \varepsilon^{12}E_2 - \varepsilon^{13}E_3\right] + \varepsilon^{32}E_2 + \varepsilon^{33}E_3 \\ &= \left(\frac{\varepsilon^{31}}{\varepsilon^{11}}\right)D^1 + \left(\varepsilon^{32} - \frac{\varepsilon^{31}\varepsilon^{12}}{\varepsilon^{11}}\right)E_2 + \left(\varepsilon^{33} - \frac{\varepsilon^{31}\varepsilon^{13}}{\varepsilon^{11}}\right)E_3\,, \end{aligned} \tag{A.7}$$

and

[1] More precisely, even every single matrix entry is a function of x^2

$$
\begin{aligned}
\tilde{\mathbf{D}}^3 &= \left[\!\!\left[\frac{\varepsilon^{31}}{\varepsilon^{11}}\right]\!\!\right]\left(\left[\!\!\left[\frac{1}{\varepsilon^{11}}\right]\!\!\right]^{-1}\tilde{\mathbf{E}}_1 + \left[\!\!\left[\frac{1}{\varepsilon^{11}}\right]\!\!\right]^{-1}\left[\!\!\left[\frac{\varepsilon^{12}}{\varepsilon^{11}}\right]\!\!\right]\tilde{\mathbf{E}}_2 + \left[\!\!\left[\frac{1}{\varepsilon^{11}}\right]\!\!\right]^{-1}\left[\!\!\left[\frac{\varepsilon^{13}}{\varepsilon^{11}}\right]\!\!\right]\tilde{\mathbf{E}}_3\right) \\
&\quad + \left[\!\!\left[\varepsilon^{32} - \frac{\varepsilon^{31}\varepsilon^{12}}{\varepsilon^{11}}\right]\!\!\right]\tilde{\mathbf{E}}_2 + \left[\!\!\left[\varepsilon^{33} - \frac{\varepsilon^{31}\varepsilon^{13}}{\varepsilon^{11}}\right]\!\!\right]\tilde{\mathbf{E}}_3 \\
&= \left[\!\!\left[\frac{\varepsilon^{31}}{\varepsilon^{11}}\right]\!\!\right]\left[\!\!\left[\frac{1}{\varepsilon^{11}}\right]\!\!\right]^{-1}\tilde{\mathbf{E}}_1 + \left(\left[\!\!\left[\varepsilon^{32} - \frac{\varepsilon^{31}\varepsilon^{12}}{\varepsilon^{11}}\right]\!\!\right] + \left[\!\!\left[\frac{\varepsilon^{31}}{\varepsilon^{11}}\right]\!\!\right]\left[\!\!\left[\frac{1}{\varepsilon^{11}}\right]\!\!\right]^{-1}\left[\!\!\left[\frac{\varepsilon^{12}}{\varepsilon^{11}}\right]\!\!\right]\right)\tilde{\mathbf{E}}_2 \\
&\quad + \left(\left[\!\!\left[\varepsilon^{33} - \frac{\varepsilon^{31}\varepsilon^{13}}{\varepsilon^{11}}\right]\!\!\right] + \left[\!\!\left[\frac{\varepsilon^{31}}{\varepsilon^{11}}\right]\!\!\right]\left[\!\!\left[\frac{1}{\varepsilon^{11}}\right]\!\!\right]^{-1}\left[\!\!\left[\frac{\varepsilon^{13}}{\varepsilon^{11}}\right]\!\!\right]\right)\tilde{\mathbf{E}}_3 \\
&=: \quad \underline{\mathbf{Q}}^{31}\,\tilde{\mathbf{E}}_1 \quad + \quad \underline{\mathbf{Q}}^{32}\,\tilde{\mathbf{E}}_2 \quad + \quad \underline{\mathbf{Q}}^{33}\,\tilde{\mathbf{E}}_3\,.
\end{aligned} \tag{A.8}
$$

In the following we will perform the Fourier transformation along this direction. Then, the fields D^2, E_1, and E_3 must be continuous across the whole unit cell.

We start with reformulating Eq. (A.6) such that we only have products of Laurent type

$$
\tilde{\mathbf{E}}_2 = \left(\underline{\mathbf{Q}}^{22}\right)^{-1}\tilde{\mathbf{D}}^2 \;-\; \left(\left(\underline{\mathbf{Q}}^{22}\right)^{-1}\underline{\mathbf{Q}}^{21}\right)\tilde{\mathbf{E}}_1 \;-\; \left(\left(\underline{\mathbf{Q}}^{22}\right)^{-1}\underline{\mathbf{Q}}^{23}\right)\tilde{\mathbf{E}}_3\,, \tag{A.9}
$$

and Fourier transform it into

$$
\begin{aligned}
\tilde{\tilde{\mathbf{E}}}_2 &= \left[\!\!\left[\left(\underline{\mathbf{Q}}^{22}\right)^{-1}\right]\!\!\right]\tilde{\tilde{\mathbf{D}}}^2 \;-\; \left[\!\!\left[\left(\underline{\mathbf{Q}}^{22}\right)^{-1}\;\underline{\mathbf{Q}}^{21}\right]\!\!\right]\tilde{\tilde{\mathbf{E}}}_1 \;-\; \left[\!\!\left[\left(\underline{\mathbf{Q}}^{22}\right)^{-1}\;\underline{\mathbf{Q}}^{23}\right]\!\!\right]\tilde{\tilde{\mathbf{E}}}_3 \\
&=: \underline{\bar{\mathbf{Q}}}^{22}\,\tilde{\tilde{\mathbf{D}}}^2 \;-\; \underline{\bar{\mathbf{Q}}}^{21}\,\tilde{\tilde{\mathbf{E}}}_1 \;-\; \underline{\bar{\mathbf{Q}}}^{23}\,\tilde{\tilde{\mathbf{E}}}_3\,.
\end{aligned} \tag{A.10}
$$

Equation (A.10) in turn can be solved for $\tilde{\tilde{\mathbf{D}}}^2$ again, which leads to

$$
\begin{aligned}
\tilde{\tilde{\mathbf{D}}}^2 &= \left(\left(\underline{\bar{\mathbf{Q}}}^{22}\right)^{-1}\underline{\bar{\mathbf{Q}}}^{21}\right)\tilde{\tilde{\mathbf{E}}}_1 \quad + \quad \left(\underline{\bar{\mathbf{Q}}}^{22}\right)^{-1}\tilde{\tilde{\mathbf{E}}}_2 \quad + \quad \left(\left(\underline{\bar{\mathbf{Q}}}^{22}\right)^{-1}\underline{\bar{\mathbf{Q}}}^{23}\right)\tilde{\tilde{\mathbf{E}}}_3 \\
&=: \quad \underline{\tilde{\varepsilon}}^{21}\,\tilde{\tilde{\mathbf{E}}}_1 \quad + \quad \underline{\tilde{\varepsilon}}^{22}\,\tilde{\tilde{\mathbf{E}}}_2 \quad + \quad \underline{\tilde{\varepsilon}}^{23}\,\tilde{\tilde{\mathbf{E}}}_3\,.
\end{aligned} \tag{A.11}
$$

A similar procedure applies to the other components:

$$
\begin{aligned}
\tilde{\mathbf{D}}^1 &= \quad \underline{\mathbf{Q}}^{11}\,\tilde{\mathbf{E}}_1 \quad + \quad \underline{\mathbf{Q}}^{12}\,\tilde{\mathbf{E}}_2 \quad + \quad \underline{\mathbf{Q}}^{13}\,\tilde{\mathbf{E}}_3 \\
&\overset{\text{Eq. (A.9)}}{=} \quad \underline{\mathbf{Q}}^{11}\,\tilde{\mathbf{E}}_1 \quad + \quad \underline{\mathbf{Q}}^{12}\left(\underline{\mathbf{Q}}^{22}\right)^{-1}\left(\tilde{\mathbf{D}}^2 - \underline{\mathbf{Q}}^{21}\tilde{\mathbf{E}}_1 - \underline{\mathbf{Q}}^{23}\tilde{\mathbf{E}}_3\right) \quad + \quad \underline{\mathbf{Q}}^{13}\,\tilde{\mathbf{E}}_3 \\
&= \quad \left(\underline{\mathbf{Q}}^{11} - \underline{\mathbf{Q}}^{12}\left(\underline{\mathbf{Q}}^{22}\right)^{-1}\underline{\mathbf{Q}}^{21}\right)\tilde{\mathbf{E}}_1 \quad + \quad \left(\underline{\mathbf{Q}}^{12}\left(\underline{\mathbf{Q}}^{22}\right)^{-1}\right)\tilde{\mathbf{D}}^2 \\
&\quad + \quad \left(\underline{\mathbf{Q}}^{13} - \underline{\mathbf{Q}}^{12}\left(\underline{\mathbf{Q}}^{22}\right)^{-1}\underline{\mathbf{Q}}^{23}\right)\tilde{\mathbf{E}}_3\,,
\end{aligned} \tag{A.12}
$$

with Fourier transform

$$\begin{aligned}
\tilde{\underline{\mathbf{D}}}^1 &= \left[\underline{\mathbf{Q}}^{11} - \underline{\mathbf{Q}}^{12}\left(\underline{\mathbf{Q}}^{22}\right)^{-1}\underline{\mathbf{Q}}^{21}\right]\tilde{\underline{\mathbf{E}}}_1 + \left[\underline{\mathbf{Q}}^{12}\left(\underline{\mathbf{Q}}^{22}\right)^{-1}\right]\tilde{\underline{\mathbf{D}}}^2 \\
&\quad + \left[\underline{\mathbf{Q}}^{13} - \underline{\mathbf{Q}}^{12}\left(\underline{\mathbf{Q}}^{22}\right)^{-1}\underline{\mathbf{Q}}^{23}\right]\tilde{\underline{\mathbf{E}}}_3 \\
&=: \bar{\underline{\mathbf{Q}}}^{11}\,\tilde{\underline{\mathbf{E}}}_1 + \bar{\underline{\mathbf{Q}}}^{12}\,\tilde{\underline{\mathbf{D}}}^2 + \bar{\underline{\mathbf{Q}}}^{13}\,\tilde{\underline{\mathbf{E}}}_3 \\
&\overset{\text{Eq. (A.11)}}{=} \left(\bar{\underline{\mathbf{Q}}}^{11} + \bar{\underline{\mathbf{Q}}}^{12}\left(\bar{\underline{\mathbf{Q}}}^{22}\right)^{-1}\bar{\underline{\mathbf{Q}}}^{21}\right)\tilde{\underline{\mathbf{E}}}_1 + \left(\bar{\underline{\mathbf{Q}}}^{12}\left(\bar{\underline{\mathbf{Q}}}^{22}\right)^{-1}\right)\tilde{\underline{\mathbf{E}}}_2 \\
&\quad + \left(\bar{\underline{\mathbf{Q}}}^{13} + \bar{\underline{\mathbf{Q}}}^{12}\left(\bar{\underline{\mathbf{Q}}}^{22}\right)^{-1}\bar{\underline{\mathbf{Q}}}^{23}\right)\tilde{\underline{\mathbf{E}}}_3 \\
&=: \tilde{\underline{\varepsilon}}^{11}\,\tilde{\underline{\mathbf{E}}}_1 + \tilde{\underline{\varepsilon}}^{12}\,\tilde{\underline{\mathbf{E}}}_2 + \tilde{\underline{\varepsilon}}^{13}\,\tilde{\underline{\mathbf{E}}}_3\,,
\end{aligned} \tag{A.13}$$

and

$$\begin{aligned}
\tilde{\mathbf{D}}^3 &= \underline{\mathbf{Q}}^{31}\,\tilde{\mathbf{E}}_1 + \underline{\mathbf{Q}}^{32}\,\tilde{\mathbf{E}}_2 + \underline{\mathbf{Q}}^{33}\,\tilde{\mathbf{E}}_3 \\
&\overset{\text{Eq. (A.9)}}{=} \underline{\mathbf{Q}}^{31}\,\tilde{\mathbf{E}}_1 + \underline{\mathbf{Q}}^{32}\left(\underline{\mathbf{Q}}^{22}\right)^{-1}\left(\tilde{\mathbf{D}}^2 - \underline{\mathbf{Q}}^{21}\tilde{\mathbf{E}}_1 - \underline{\mathbf{Q}}^{23}\tilde{\mathbf{E}}_3\right) + \underline{\mathbf{Q}}^{13}\,\tilde{\mathbf{E}}_3 \\
&= \left(\underline{\mathbf{Q}}^{31} - \underline{\mathbf{Q}}^{32}\left(\underline{\mathbf{Q}}^{22}\right)^{-1}\underline{\mathbf{Q}}^{21}\right)\tilde{\mathbf{E}}_1 + \left(\underline{\mathbf{Q}}^{32}\left(\underline{\mathbf{Q}}^{22}\right)^{-1}\right)\tilde{\mathbf{D}}^2 \\
&\quad + \left(\underline{\mathbf{Q}}^{33} - \underline{\mathbf{Q}}^{32}\left(\underline{\mathbf{Q}}^{22}\right)^{-1}\underline{\mathbf{Q}}^{23}\right)\tilde{\mathbf{E}}_3\,,
\end{aligned} \tag{A.14}$$

with Fourier transform

$$\begin{aligned}
\tilde{\underline{\mathbf{D}}}^3 &= \left[\underline{\mathbf{Q}}^{31} - \underline{\mathbf{Q}}^{32}\left(\underline{\mathbf{Q}}^{22}\right)^{-1}\underline{\mathbf{Q}}^{21}\right]\tilde{\underline{\mathbf{E}}}_1 + \left[\underline{\mathbf{Q}}^{32}\left(\underline{\mathbf{Q}}^{22}\right)^{-1}\right]\tilde{\underline{\mathbf{D}}}^2 \\
&\quad + \left[\underline{\mathbf{Q}}^{33} - \underline{\mathbf{Q}}^{32}\left(\underline{\mathbf{Q}}^{22}\right)^{-1}\underline{\mathbf{Q}}^{23}\right]\tilde{\underline{\mathbf{E}}}_3 \\
&=: \bar{\underline{\mathbf{Q}}}^{31}\,\tilde{\underline{\mathbf{E}}}_1 + \bar{\underline{\mathbf{Q}}}^{32}\,\tilde{\underline{\mathbf{D}}}^2 + \bar{\underline{\mathbf{Q}}}^{33}\,\tilde{\underline{\mathbf{E}}}_3 \\
&\overset{\text{Eq. (A.11)}}{=} \left(\bar{\underline{\mathbf{Q}}}^{31} + \bar{\underline{\mathbf{Q}}}^{32}\left(\bar{\underline{\mathbf{Q}}}^{22}\right)^{-1}\bar{\underline{\mathbf{Q}}}^{21}\right)\tilde{\underline{\mathbf{E}}}_1 + \left(\bar{\underline{\mathbf{Q}}}^{32}\left(\bar{\underline{\mathbf{Q}}}^{22}\right)^{-1}\right)\tilde{\underline{\mathbf{E}}}_2 \\
&\quad + \left(\bar{\underline{\mathbf{Q}}}^{33} + \bar{\underline{\mathbf{Q}}}^{32}\left(\bar{\underline{\mathbf{Q}}}^{22}\right)^{-1}\bar{\underline{\mathbf{Q}}}^{23}\right)\tilde{\underline{\mathbf{E}}}_3 \\
&=: \tilde{\underline{\varepsilon}}^{31}\,\tilde{\underline{\mathbf{E}}}_1 + \tilde{\underline{\varepsilon}}^{32}\,\tilde{\underline{\mathbf{E}}}_2 + \tilde{\underline{\varepsilon}}^{33}\,\tilde{\underline{\mathbf{E}}}_3\,.
\end{aligned} \tag{A.15}$$

B Appendix B.
Eigenvalue Problems

B.1. Full Anisotropic Eigenvalue Equation

The covariant formulation of Maxwell's equations has been presented in Sec. 2.5.1. Maxwell's equations, Eqs. (2.5.3), written component wise, read

$$\frac{\partial_2}{\mathrm{i}} E_3 - \frac{\partial_3}{\mathrm{i}} E_2 = \omega^2 \left(\mu^{11} H_1 + \mu^{12} H_2 + \mu^{13} H_3 \right) , \tag{B.1a}$$

$$\frac{\partial_3}{\mathrm{i}} E_1 - \frac{\partial_1}{\mathrm{i}} E_3 = \omega^2 \left(\mu^{21} H_1 + \mu^{22} H_2 + \mu^{23} H_3 \right) , \tag{B.1b}$$

$$\frac{\partial_1}{\mathrm{i}} E_2 - \frac{\partial_2}{\mathrm{i}} E_1 = \omega^2 \left(\mu^{31} H_1 + \mu^{32} H_2 + \mu^{33} H_3 \right) , \tag{B.1c}$$

and

$$\frac{\partial_2}{\mathrm{i}} H_3 - \frac{\partial_3}{\mathrm{i}} H_2 = - \left(\varepsilon^{11} E_1 + \varepsilon^{12} E_2 + \varepsilon^{13} E_3 \right) , \tag{B.2a}$$

$$\frac{\partial_3}{\mathrm{i}} H_1 - \frac{\partial_1}{\mathrm{i}} H_3 = - \left(\varepsilon^{21} E_1 + \varepsilon^{22} E_2 + \varepsilon^{23} E_3 \right) , \tag{B.2b}$$

$$\frac{\partial_1}{\mathrm{i}} H_2 - \frac{\partial_2}{\mathrm{i}} H_1 = - \left(\varepsilon^{31} E_1 + \varepsilon^{32} E_2 + \varepsilon^{33} E_3 \right) . \tag{B.2c}$$

In order to derive the associated *full ansotropic eigenvalue equation*, we solve

Eq. (B.1a)		$\frac{\partial_3}{\mathrm{i}} \left(-E_2 \right)$,
Eq. (B.1b)		$\frac{\partial_3}{\mathrm{i}} E_1$,
Eq. (B.1c)	for	H_3 ,
Eq. (B.2a)		$\frac{\partial_3}{\mathrm{i}} H_2$,
Eq. (B.2b)		$\frac{\partial_3}{\mathrm{i}} H_1$,
Eq. (B.2c)		E_3 .

Then, we get

$$\frac{\partial_3}{\mathrm{i}}\left(-E_2\right) = \omega^2\left(\mu^{11}H_1 + \mu^{12}H_2 + \mu^{13}H_3\right) - \frac{\partial_2}{\mathrm{i}}E_3, \tag{B.3a}$$

$$\frac{\partial_3}{\mathrm{i}}E_1 = \omega^2\left(\mu^{21}H_1 + \mu^{22}H_2 + \mu^{23}H_3\right) + \frac{\partial_1}{\mathrm{i}}E_3, \tag{B.3b}$$

$$H_3 = (\mu^{33})^{-1}\left(\frac{1}{\omega^2}\Big(-\frac{\partial_1}{\mathrm{i}}\left(-E_2\right) - \frac{\partial_2}{\mathrm{i}}E_1\Big) - \mu^{31}H_1 - \mu^{32}H_2\right) \tag{B.3c}$$

and

$$\frac{\partial_3}{\mathrm{i}}H_2 = +\left(\varepsilon^{11}E_1 + \varepsilon^{12}E_2 + \varepsilon^{13}E_3\right) + \frac{\partial_2}{\mathrm{i}}H_3, \tag{B.4a}$$

$$\frac{\partial_3}{\mathrm{i}}H_1 = -\left(\varepsilon^{21}E_1 + \varepsilon^{22}E_2 + \varepsilon^{23}E_3\right) + \frac{\partial_1}{\mathrm{i}}H_3, \tag{B.4b}$$

$$E_3 = (\varepsilon^{33})^{-1}\left(\Big(\frac{\partial_2}{\mathrm{i}}H_1 - \frac{\partial_1}{\mathrm{i}}H_2\Big) - \varepsilon^{31}E_1 + \varepsilon^{32}\left(-E_2\right)\right). \tag{B.4c}$$

We now substitute B.3c and B.4c into B.3a

$$\begin{aligned}
\frac{\partial_3}{\mathrm{i}}\left(-E_2\right) &= \omega^2\left(\mu^{11}H_1 + \mu^{12}H_2 + \mu^{13}H_3\right) - \frac{\partial_2}{\mathrm{i}}E_3 \\
&= \omega^2\left(\mu^{11}H_1 + \mu^{12}H_2 + \mu^{13}\left((\mu^{33})^{-1}\left(\frac{1}{\omega^2}\Big(-\frac{\partial_1}{\mathrm{i}}\left(-E_2\right) - \frac{\partial_2}{\mathrm{i}}E_1\Big) - \mu^{31}H_1 - \mu^{32}H_2\right)\right)\right) \\
&\quad - \frac{\partial_2}{\mathrm{i}}\left((\varepsilon^{33})^{-1}\left(\Big(\frac{\partial_2}{\mathrm{i}}H_1 - \frac{\partial_1}{\mathrm{i}}H_2\Big) - \varepsilon^{31}E_1 + \varepsilon^{32}\left(-E_2\right)\right)\right) \\
&= \omega^2\left(\mu^{11} - \mu^{13}(\mu^{33})^{-1}\mu^{31}\right)H_1 + \omega^2\left(\mu^{12} - \mu^{13}(\mu^{33})^{-1}\mu^{32}\right)H_2 \\
&\quad - \omega^2\left(\mu^{13}(\mu^{33})^{-1}\frac{1}{\omega^2}\frac{\partial_1}{\mathrm{i}}\right)\left(-E_2\right) - \omega^2\left(\mu^{13}(\mu^{33})^{-1}\frac{1}{\omega^2}\frac{\partial_2}{\mathrm{i}}\right)E_1 \\
&\quad - \left(\frac{\partial_2}{\mathrm{i}}(\varepsilon^{33})^{-1}\frac{\partial_2}{\mathrm{i}}\right)H_1 + \left(\frac{\partial_2}{\mathrm{i}}(\varepsilon^{33})^{-1}\frac{\partial_1}{\mathrm{i}}\right)H_2 \\
&\quad + \left(\frac{\partial_2}{\mathrm{i}}(\varepsilon^{33})^{-1}\varepsilon^{31}\right)E_1 - \left(\frac{\partial_2}{\mathrm{i}}(\varepsilon^{33})^{-1}\varepsilon^{32}\right)\left(-E_2\right) \\
&= \Big(-\frac{\partial_2}{\mathrm{i}}(\varepsilon^{33})^{-1}\varepsilon^{32} - \mu^{13}(\mu^{33})^{-1}\frac{\partial_1}{\mathrm{i}}\Big)\left(-E_2\right) \\
&\quad + \Big(\frac{\partial_2}{\mathrm{i}}(\varepsilon^{33})^{-1}\varepsilon^{31} - \mu^{13}(\mu^{33})^{-1}\frac{\partial_2}{\mathrm{i}}\Big)E_1 \\
&\quad + \Big(\omega^2\left(\mu^{11} - \mu^{13}(\mu^{33})^{-1}\mu^{31}\right) - \frac{\partial_2}{\mathrm{i}}(\varepsilon^{33})^{-1}\frac{\partial_2}{\mathrm{i}}\Big)H_1 \\
&\quad + \Big(\omega^2\left(\mu^{12} - \mu^{13}(\mu^{33})^{-1}\mu^{32}\right) + \frac{\partial_2}{\mathrm{i}}(\varepsilon^{33})^{-1}\frac{\partial_1}{\mathrm{i}}\Big)H_2
\end{aligned} \tag{B.5}$$

and also into B.3b

$$
\begin{aligned}
\frac{\partial_3}{\mathrm{i}} E_1 &= \omega^2 \left(\mu^{21} H_1 + \mu^{22} H_2 + \mu^{23} H_3\right) + \frac{\partial_1}{\mathrm{i}} E_3 \\
&= \omega^2 \left(\mu^{21} H_1 + \mu^{22} H_2 + \mu^{23} (\mu^{33})^{-1} \left(\frac{1}{\omega^2}\left(-\frac{\partial_1}{\mathrm{i}}\left(-E_2\right) - \frac{\partial_2}{\mathrm{i}} E_1\right) - \mu^{31} H_1 - \mu^{32} H_2\right)\right) \\
&\quad + \frac{\partial_1}{\mathrm{i}} (\varepsilon^{33})^{-1} \left(\left(\frac{\partial_2}{\mathrm{i}} H_1 - \frac{\partial_1}{\mathrm{i}} H_2\right) - \varepsilon^{31} E_1 + \varepsilon^{32}\left(-E_2\right)\right) \\
&= \omega^2 \left(\mu^{21} - \mu^{23} (\mu^{33})^{-1} \mu^{31}\right) H_1 + \omega^2 \left(\mu^{22} - \mu^{23} (\mu^{33})^{-1} \mu^{32}\right) H_2 \\
&\quad - \left(\mu^{23} (\mu^{33})^{-1} \frac{\partial_1}{\mathrm{i}}\right)\left(-E_2\right) - \left(\mu^{23} (\mu^{33})^{-1} \frac{\partial_2}{\mathrm{i}}\right) E_1 \\
&\quad + \left(\frac{\partial_1}{\mathrm{i}} (\varepsilon^{33})^{-1} \frac{\partial_2}{\mathrm{i}}\right) H_1 - \left(\frac{\partial_1}{\mathrm{i}} (\varepsilon^{33})^{-1} \frac{\partial_1}{\mathrm{i}}\right) H_2 \\
&\quad - \left(\frac{\partial_1}{\mathrm{i}} (\varepsilon^{33})^{-1} \varepsilon^{31}\right) E_1 + \left(\frac{\partial_1}{\mathrm{i}} (\varepsilon^{33})^{-1} \varepsilon^{32}\right)\left(-E_2\right) \\
&= \left(\frac{\partial_1}{\mathrm{i}} (\varepsilon^{33})^{-1} \varepsilon^{32} - \mu^{23} (\mu^{33})^{-1} \frac{\partial_1}{\mathrm{i}}\right)\left(-E_2\right) \\
&\quad + \left(-\frac{\partial_1}{\mathrm{i}} (\varepsilon^{33})^{-1} \varepsilon^{31} - \mu^{23} (\mu^{33})^{-1} \frac{\partial_2}{\mathrm{i}}\right) E_1 \\
&\quad + \left(\omega^2 \left(\mu^{21} - \mu^{23} (\mu^{33})^{-1} \mu^{31}\right) + \frac{\partial_1}{\mathrm{i}} (\varepsilon^{33})^{-1} \frac{\partial_2}{\mathrm{i}}\right) H_1 \\
&\quad + \left(\omega^2 \left(\mu^{22} - \mu^{23} (\mu^{33})^{-1} \mu^{32}\right) - \frac{\partial_1}{\mathrm{i}} (\varepsilon^{33})^{-1} \frac{\partial_1}{\mathrm{i}}\right) H_2 \,. \qquad \text{(B.6)}
\end{aligned}
$$

We repeat the same substitutions for B.4b

$$
\begin{aligned}
\frac{\partial_3}{\mathrm{i}} H_1 &= -\left(\varepsilon^{21} E_1 + \varepsilon^{22} E_2 + \varepsilon^{23} E_3\right) + \frac{\partial_1}{\mathrm{i}} H_3 \\
&= -\left(\varepsilon^{21} E_1 + \varepsilon^{22} E_2 + \varepsilon^{23} (\varepsilon^{33})^{-1} \left(\left(\frac{\partial_2}{\mathrm{i}} H_1 - \frac{\partial_1}{\mathrm{i}} H_2\right) - \varepsilon^{31} E_1 + \varepsilon^{32}\left(-E_2\right)\right)\right) \\
&\quad + \frac{\partial_1}{\mathrm{i}} (\mu^{33})^{-1} \left(\frac{1}{\omega^2}\left(-\frac{\partial_1}{\mathrm{i}}\left(-E_2\right) - \frac{\partial_2}{\mathrm{i}} E_1\right) - \mu^{31} H_1 - \mu^{32} H_2\right) \\
&= \left(\varepsilon^{22} - \varepsilon^{23} (\varepsilon^{33})^{-1} \varepsilon^{32}\right)\left(-E_2\right) - \left(\varepsilon^{21} - \varepsilon^{23} (\varepsilon^{33})^{-1} \varepsilon^{31}\right) E_1 \\
&\quad - \left(\varepsilon^{23} (\varepsilon^{33})^{-1} \frac{\partial_2}{\mathrm{i}}\right) H_1 + \left(\varepsilon^{23} (\varepsilon^{33})^{-1} \frac{\partial_1}{\mathrm{i}}\right) H_2 \\
&\quad - \left(\frac{1}{\omega^2} \frac{\partial_1}{\mathrm{i}} (\mu^{33})^{-1} \frac{\partial_1}{\mathrm{i}}\right)\left(-E_2\right) - \left(\frac{1}{\omega^2} \frac{\partial_1}{\mathrm{i}} (\mu^{33})^{-1} \frac{\partial_2}{\mathrm{i}}\right) E_1 \\
&\quad - \left(\frac{\partial_1}{\mathrm{i}} (\mu^{33})^{-1} \mu^{31}\right) H_1 - \left(\frac{\partial_1}{\mathrm{i}} (\mu^{33})^{-1} \mu^{32}\right) H_2
\end{aligned}
$$

$$
\begin{aligned}
&= \left(\left(\varepsilon^{22} - \varepsilon^{23}(\varepsilon^{33})^{-1}\varepsilon^{32}\right) - \frac{1}{\omega^2}\frac{\partial_1}{\mathrm{i}}(\mu^{33})^{-1}\frac{\partial_1}{\mathrm{i}} \right)(-E_2) \\
&\quad + \left(- \left(\varepsilon^{21} - \varepsilon^{23}(\varepsilon^{33})^{-1}\varepsilon^{31}\right) - \frac{1}{\omega^2}\frac{\partial_1}{\mathrm{i}}(\mu^{33})^{-1}\frac{\partial_2}{\mathrm{i}} \right) E_1 \\
&\quad + \left(- \left(\varepsilon^{23}(\varepsilon^{33})^{-1}\right)\frac{\partial_2}{\mathrm{i}} - \frac{\partial_1}{\mathrm{i}}\left((\mu^{33})^{-1}\mu^{31}\right) \right) H_1 \\
&\quad + \left(\left(\varepsilon^{23}(\varepsilon^{33})^{-1}\right)\frac{\partial_1}{\mathrm{i}} - \frac{\partial_1}{\mathrm{i}}\left((\mu^{33})^{-1}\mu^{32}\right) \right) H_2
\end{aligned} \tag{B.7}
$$

as well as for B.4a

$$
\begin{aligned}
\frac{\partial_3}{\mathrm{i}}H_2 &= + \left(\varepsilon^{11}E_1 + \varepsilon^{12}E_2 + \varepsilon^{13}E_3\right) + \frac{\partial_2}{\mathrm{i}}H_3 \\
&= + \left(\varepsilon^{11}E_1 + \varepsilon^{12}E_2 + \varepsilon^{13}(\varepsilon^{33})^{-1}\left(\left(\frac{\partial_2}{\mathrm{i}}H_1 - \frac{\partial_1}{\mathrm{i}}H_2\right) - \varepsilon^{31}E_1 + \varepsilon^{32}\left(-E_2\right) \right) \right) \\
&\quad + \frac{\partial_2}{\mathrm{i}}(\mu^{33})^{-1}\left(\frac{1}{\omega^2}\left(- \frac{\partial_1}{\mathrm{i}}\left(-E_2\right) - \frac{\partial_2}{\mathrm{i}}E_1 \right) - \mu^{31}H_1 - \mu^{32}H_2 \right) \\
&= - \left(\varepsilon^{12} - \varepsilon^{13}(\varepsilon^{33})^{-1}\varepsilon^{32}\right)\left(-E_2\right) + \left(\varepsilon^{11} - \varepsilon^{13}(\varepsilon^{33})^{-1}\varepsilon^{31}\right)E_1 \\
&\quad + \left(\varepsilon^{13}(\varepsilon^{33})^{-1}\frac{\partial_2}{\mathrm{i}} \right) H_1 - \left(\varepsilon^{13}(\varepsilon^{33})^{-1}\frac{\partial_1}{\mathrm{i}} \right) H_2 \\
&\quad - \left(\frac{1}{\omega^2}\frac{\partial_2}{\mathrm{i}}(\mu^{33})^{-1}\frac{\partial_1}{\mathrm{i}} \right)\left(-E_2\right) - \left(\frac{1}{\omega^2}\frac{\partial_2}{\mathrm{i}}(\mu^{33})^{-1}\frac{\partial_2}{\mathrm{i}} \right) E_1 \\
&\quad - \left(\frac{\partial_2}{\mathrm{i}}(\mu^{33})^{-1}\mu^{31} \right) H_1 - \left(\frac{\partial_2}{\mathrm{i}}(\mu^{33})^{-1}\mu^{32} \right) H_2 \\
&= \left(- \left(\varepsilon^{12} - \varepsilon^{13}(\varepsilon^{33})^{-1}\varepsilon^{32}\right) - \frac{1}{\omega^2}\frac{\partial_2}{\mathrm{i}}(\mu^{33})^{-1}\frac{\partial_1}{\mathrm{i}} \right)(-E_2) \\
&\quad + \left(\left(\varepsilon^{11} - \varepsilon^{13}(\varepsilon^{33})^{-1}\varepsilon^{31}\right) - \frac{1}{\omega^2}\frac{\partial_2}{\mathrm{i}}(\mu^{33})^{-1}\frac{\partial_2}{\mathrm{i}} \right) E_1 \\
&\quad + \left(\left(\varepsilon^{13}(\varepsilon^{33})^{-1}\right)\frac{\partial_2}{\mathrm{i}} - \frac{\partial_2}{\mathrm{i}}\left((\mu^{33})^{-1}\mu^{31}\right) \right) H_1 \\
&\quad + \left(- \left(\varepsilon^{13}(\varepsilon^{33})^{-1}\right)\frac{\partial_1}{\mathrm{i}} - \frac{\partial_2}{\mathrm{i}}\left((\mu^{33})^{-1}\mu^{32}\right) \right) H_2
\end{aligned} \tag{B.8}
$$

In order to improve readability, the composed terms of matrix elements for the permittivity and permeability above can be merged into new quantities $\underline{\breve{\boldsymbol{\varepsilon}}} = \hat{l}_3^-(\underline{\boldsymbol{\varepsilon}})$ and $\underline{\breve{\boldsymbol{\mu}}} = \hat{l}_3^-(\underline{\boldsymbol{\mu}})$, whose components read

$$\breve{\varepsilon}^{11} = \varepsilon^{11} - \varepsilon^{13}(\varepsilon^{33})^{-1}\varepsilon^{31} \tag{B.9a}$$

$$\breve{\varepsilon}^{12} = \varepsilon^{12} - \varepsilon^{13}(\varepsilon^{33})^{-1}\varepsilon^{32} \tag{B.9b}$$

$$\breve{\varepsilon}^{13} = \varepsilon^{13}(\varepsilon^{33})^{-1} \tag{B.9c}$$

$$\breve{\varepsilon}^{21} = \varepsilon^{21} - \varepsilon^{23}(\varepsilon^{33})^{-1}\varepsilon^{31} \tag{B.9d}$$
$$\breve{\varepsilon}^{22} = \varepsilon^{22} - \varepsilon^{23}(\varepsilon^{33})^{-1}\varepsilon^{32} \tag{B.9e}$$
$$\breve{\varepsilon}^{23} = \varepsilon^{23}(\varepsilon^{33})^{-1} \tag{B.9f}$$
$$\breve{\varepsilon}^{31} = (\varepsilon^{33})^{-1}\varepsilon^{31} \tag{B.9g}$$
$$\breve{\varepsilon}^{32} = (\varepsilon^{33})^{-1}\varepsilon^{32} \tag{B.9h}$$
$$\breve{\varepsilon}^{33} = (\varepsilon^{33})^{-1}\ , \tag{B.9i}$$

and

$$\breve{\mu}^{11} = \mu^{11} - \mu^{13}(\mu^{33})^{-1}\mu^{31} \tag{B.10a}$$
$$\breve{\mu}^{12} = \mu^{12} - \mu^{13}(\mu^{33})^{-1}\mu^{32} \tag{B.10b}$$
$$\breve{\mu}^{13} = \mu^{13}(\mu^{33})^{-1} \tag{B.10c}$$
$$\breve{\mu}^{21} = \mu^{21} - \mu^{23}(\mu^{33})^{-1}\mu^{31} \tag{B.10d}$$
$$\breve{\mu}^{22} = \mu^{22} - \mu^{23}(\mu^{33})^{-1}\mu^{32} \tag{B.10e}$$
$$\breve{\mu}^{23} = \mu^{23}(\mu^{33})^{-1} \tag{B.10f}$$
$$\breve{\mu}^{31} = (\mu^{33})^{-1}\mu^{31} \tag{B.10g}$$
$$\breve{\mu}^{32} = (\mu^{33})^{-1}\mu^{32} \tag{B.10h}$$
$$\breve{\mu}^{33} = (\mu^{33})^{-1}\ . \tag{B.10i}$$

The used operator is Li's operator $\hat{l}_3^-$ stated in Ref. [51]. This replacement leads to simplifieded versions of B.5, B.6, B.7, and B.8

$$\begin{aligned}\frac{\partial_3}{\mathrm{i}}(-E_2) = &\left(-\frac{\partial_2}{\mathrm{i}}\breve{\varepsilon}^{32} - \breve{\mu}^{13}\frac{\partial_1}{\mathrm{i}}\right)(-E_2) \quad + \quad \left(\frac{\partial_2}{\mathrm{i}}\breve{\varepsilon}^{31} - \breve{\mu}^{13}\frac{\partial_2}{\mathrm{i}}\right)E_1 \\ &+ \left(\omega^2\breve{\mu}^{11} - \frac{\partial_2}{\mathrm{i}}\breve{\varepsilon}^{33}\frac{\partial_2}{\mathrm{i}}\right)H_1 \quad + \quad \left(\omega^2\breve{\mu}^{12} + \frac{\partial_2}{\mathrm{i}}\breve{\varepsilon}^{33}\frac{\partial_1}{\mathrm{i}}\right)H_2\ ,\end{aligned} \tag{B.11}$$

$$\begin{aligned}\frac{\partial_3}{\mathrm{i}}E_1 = &\left(\frac{\partial_1}{\mathrm{i}}\breve{\varepsilon}^{32} - \breve{\mu}^{23}\frac{\partial_1}{\mathrm{i}}\right)(-E_2) \quad + \quad \left(-\frac{\partial_1}{\mathrm{i}}\breve{\varepsilon}^{31} - \breve{\mu}^{23}\frac{\partial_2}{\mathrm{i}}\right)E_1 \\ &+ \left(\omega^2\breve{\mu}^{21} + \frac{\partial_1}{\mathrm{i}}\breve{\varepsilon}^{33}\frac{\partial_2}{\mathrm{i}}\right)H_1 \quad + \quad \left(\omega^2\breve{\mu}^{22} - \frac{\partial_1}{\mathrm{i}}\breve{\varepsilon}^{33}\frac{\partial_1}{\mathrm{i}}\right)H_2\ ,\end{aligned} \tag{B.12}$$

$$\begin{aligned}\frac{\partial_3}{\mathrm{i}}H_1 = &\left(\breve{\varepsilon}^{22} - \frac{1}{\omega^2}\frac{\partial_1}{\mathrm{i}}\breve{\mu}^{33}\frac{\partial_1}{\mathrm{i}}\right)(-E_2) \quad + \quad \left(-\breve{\varepsilon}^{21} - \frac{1}{\omega^2}\frac{\partial_1}{\mathrm{i}}\breve{\mu}^{33}\frac{\partial_2}{\mathrm{i}}\right)E_1 \\ &+ \left(-\breve{\varepsilon}^{23}\frac{\partial_2}{\mathrm{i}} - \frac{\partial_1}{\mathrm{i}}\breve{\mu}^{31}\right)H_1 \quad + \quad \left(\breve{\varepsilon}^{23}\frac{\partial_1}{\mathrm{i}} - \frac{\partial_1}{\mathrm{i}}\breve{\mu}^{32}\right)H_2\ ,\end{aligned} \tag{B.13}$$

$$\begin{aligned}\frac{\partial_3}{\mathrm{i}}H_2 = &\left(-\breve{\varepsilon}^{12} - \frac{1}{\omega^2}\frac{\partial_2}{\mathrm{i}}\breve{\mu}^{33}\frac{\partial_1}{\mathrm{i}}\right)(-E_2) \quad + \quad \left(\breve{\varepsilon}^{11} - \frac{1}{\omega^2}\frac{\partial_2}{\mathrm{i}}\breve{\mu}^{33}\frac{\partial_2}{\mathrm{i}}\right)E_1 \\ &+ \left(\breve{\varepsilon}^{13}\frac{\partial_2}{\mathrm{i}} - \frac{\partial_2}{\mathrm{i}}\breve{\mu}^{31}\right)H_1 \quad + \quad \left(-\breve{\varepsilon}^{13}\frac{\partial_1}{\mathrm{i}} - \frac{\partial_2}{\mathrm{i}}\breve{\mu}^{32}\right)H_2\ .\end{aligned} \tag{B.14}$$

The z-components B.3c and B.4c similarly simplyfy to

$$H_3 = \breve{\mu}^{33}\frac{1}{\omega^2}\left(-\frac{\partial_1}{\mathrm{i}}\left(-E_2\right) - \frac{\partial_2}{\mathrm{i}}E_1\right) - \breve{\mu}^{31}H_1 - \breve{\mu}^{32}H_2\,, \tag{B.15}$$

$$E_3 = \breve{\varepsilon}^{33}\left(\frac{\partial_2}{\mathrm{i}}H_1 - \frac{\partial_1}{\mathrm{i}}H_2\right) - \breve{\varepsilon}^{31}E_1 + \breve{\varepsilon}^{32}\left(-E_2\right)\,. \tag{B.16}$$

Introducing the electromagnetic field component vector $(-E_2, E_1, H_1, H_2)^T$, and using the ansatz $e^{\mathrm{i}\gamma x^3}$ for the fields in x^3-direction which allows for the substitution $\partial_3 \to \mathrm{i}\gamma$, we combine Eq. (B.11), Eq. (B.12), Eq. (B.13), and Eq. (B.14) into a matrix equation. Thus, we finally obtain the large eigenvalue equation

$$\gamma \begin{pmatrix} -\tilde{E}_2 \\ \tilde{E}_1 \\ \tilde{H}_1 \\ \tilde{H}_2 \end{pmatrix} = \underline{\boldsymbol{\mathcal{A}}} \begin{pmatrix} -\tilde{E}_2 \\ \tilde{E}_1 \\ \tilde{H}_1 \\ \tilde{H}_2 \end{pmatrix}\,, \tag{B.17}$$

with system matrix

$$\underline{\boldsymbol{\mathcal{A}}} = \begin{pmatrix} -\frac{\partial_2}{\mathrm{i}}\breve{\varepsilon}^{32} - \breve{\mu}^{13}\frac{\partial_1}{\mathrm{i}} & \frac{\partial_2}{\mathrm{i}}\breve{\varepsilon}^{31} - \breve{\mu}^{13}\frac{\partial_2}{\mathrm{i}} & \omega^2\breve{\mu}^{11} - \frac{\partial_2}{\mathrm{i}}\breve{\varepsilon}^{33}\frac{\partial_2}{\mathrm{i}} & \omega^2\breve{\mu}^{12} + \frac{\partial_2}{\mathrm{i}}\breve{\varepsilon}^{33}\frac{\partial_1}{\mathrm{i}} \\ \frac{\partial_1}{\mathrm{i}}\breve{\varepsilon}^{32} - \breve{\mu}^{23}\frac{\partial_1}{\mathrm{i}} & -\frac{\partial_1}{\mathrm{i}}\breve{\varepsilon}^{31} - \breve{\mu}^{23}\frac{\partial_2}{\mathrm{i}} & \omega^2\breve{\mu}^{21} + \frac{\partial_1}{\mathrm{i}}\breve{\varepsilon}^{33}\frac{\partial_2}{\mathrm{i}} & \omega^2\breve{\mu}^{22} - \frac{\partial_1}{\mathrm{i}}\breve{\varepsilon}^{33}\frac{\partial_1}{\mathrm{i}} \\ \breve{\varepsilon}^{22} - \frac{1}{\omega^2}\frac{\partial_1}{\mathrm{i}}\breve{\mu}^{33}\frac{\partial_1}{\mathrm{i}} & -\breve{\varepsilon}^{21} - \frac{1}{\omega^2}\frac{\partial_1}{\mathrm{i}}\breve{\mu}^{33}\frac{\partial_2}{\mathrm{i}} & -\breve{\varepsilon}^{23}\frac{\partial_2}{\mathrm{i}} - \frac{\partial_1}{\mathrm{i}}\breve{\mu}^{31} & \breve{\varepsilon}^{23}\frac{\partial_1}{\mathrm{i}} - \frac{\partial_1}{\mathrm{i}}\breve{\mu}^{32} \\ -\breve{\varepsilon}^{12} - \frac{1}{\omega^2}\frac{\partial_2}{\mathrm{i}}\breve{\mu}^{33}\frac{\partial_1}{\mathrm{i}} & \breve{\varepsilon}^{11} - \frac{1}{\omega^2}\frac{\partial_2}{\mathrm{i}}\breve{\mu}^{33}\frac{\partial_2}{\mathrm{i}} & \breve{\varepsilon}^{13}\frac{\partial_2}{\mathrm{i}} - \frac{\partial_2}{\mathrm{i}}\breve{\mu}^{31} & -\breve{\varepsilon}^{13}\frac{\partial_1}{\mathrm{i}} - \frac{\partial_2}{\mathrm{i}}\breve{\mu}^{32} \end{pmatrix}\,. \tag{B.18}$$

B.2. smallEigenproblem

Original FMM implementation of the eigenproblem including oblique lattices.

$$\underline{\mathbf{F}} = \begin{pmatrix} \omega^2 - \underline{\beta}\left[\!\left[\epsilon\right]\!\right]^{-1}\underline{\beta} & -\omega^2\sin(\xi) + \underline{\beta}\left[\!\left[\epsilon\right]\!\right]^{-1}\underline{\alpha} \\ -\omega^2\sin(\xi) + \underline{\alpha}\left[\!\left[\epsilon\right]\!\right]^{-1}\underline{\beta} & \omega^2 - \underline{\alpha}\left[\!\left[\epsilon\right]\!\right]^{-1}\underline{\alpha} \end{pmatrix} \tag{B.19}$$

$$\underline{\mathbf{G}} = \begin{pmatrix} \left(\cos^2(\xi)\left\lceil\!\left\lfloor\epsilon\right\rfloor\!\right\rceil + \sin^2(\xi)\left[\!\left[\frac{1}{\epsilon}\right]\!\right]^{-1}\right) - \frac{1}{\omega^2}\,\underline{\alpha}\,\underline{\alpha} & \sin(\xi)\left[\!\left[\frac{1}{\epsilon}\right]\!\right]^{-1} - \frac{1}{\omega^2}\,\underline{\alpha}\,\underline{\beta} \\ \sin(\xi)\left[\!\left[\frac{1}{\epsilon}\right]\!\right]^{-1} - \frac{1}{\omega^2}\,\underline{\beta}\,\underline{\alpha} & \left(\cos^2(\xi)\left\lfloor\!\left\lceil\epsilon\right\rceil\!\right\rfloor + \sin^2(\xi)\left[\!\left[\frac{1}{\epsilon}\right]\!\right]^{-1}\right) - \frac{1}{\omega^2}\,\underline{\beta}\,\underline{\beta} \end{pmatrix} \tag{B.20}$$

$$\underline{\mathbf{eig}} = \frac{1}{\cos^2(\xi)}\,\underline{\mathbf{F}}\,\underline{\mathbf{G}}\,. \tag{B.21}$$

B.3. smallEigenproblemAdaptive

Implementation of the eigenproblem using the structure-transform real-space strategy for materials with in-plane anisotropy as necessary for AC and ASR.

$$\underline{\mathbf{F}} = \begin{pmatrix} \omega^2\,\mu^{11} - \underline{\beta}\,\varepsilon^{33}\,\underline{\beta} & \omega^2\,\mu^{12} + \underline{\beta}\,\varepsilon^{33}\,\underline{\alpha} \\ \omega^2\,\mu^{21} + \underline{\alpha}\,\varepsilon^{33}\,\underline{\beta} & \omega^2\,\mu^{22} - \underline{\alpha}\,\varepsilon^{33}\,\underline{\alpha} \end{pmatrix} \tag{B.22}$$

$$\underline{\mathbf{G}} = \begin{pmatrix} \varepsilon^{22} - \frac{1}{\omega^2}\,\underline{\alpha}\,\mu^{33}\,\underline{\alpha} & -\varepsilon^{21} - \frac{1}{\omega^2}\,\underline{\alpha}\,\mu^{33}\,\underline{\beta} \\ -\varepsilon^{12} - \frac{1}{\omega^2}\,\underline{\beta}\,\mu^{33}\,\underline{\alpha} & \varepsilon^{11} - \frac{1}{\omega^2}\,\underline{\beta}\,\mu^{33}\,\underline{\beta} \end{pmatrix} \tag{B.23}$$

B.4. smallEigenproblemPMLsimple

Implementation of PMLs for isotropic material systems with the equation-transform k-space strategy. This is a 3D version of [99].

$$\underline{\mathbf{F}} = \begin{pmatrix} \omega^2 - \left[\!\left[f_y \right]\!\right] \underline{\beta} \left[\!\left[\epsilon \right]\!\right]^{-1} \left[\!\left[f_y \right]\!\right] \underline{\beta} & \left[\!\left[f_y \right]\!\right] \underline{\beta} \left[\!\left[\epsilon \right]\!\right]^{-1} \left[\!\left[f_x \right]\!\right] \underline{\alpha} \\ \left[\!\left[f_x \right]\!\right] \underline{\alpha} \left[\!\left[\epsilon \right]\!\right]^{-1} \left[\!\left[f_y \right]\!\right] \underline{\beta} & \omega^2 - \left[\!\left[f_x \right]\!\right] \underline{\alpha} \left[\!\left[\epsilon \right]\!\right]^{-1} \left[\!\left[f_x \right]\!\right] \underline{\alpha} \end{pmatrix} \tag{B.24}$$

$$\underline{\mathbf{G}} = \begin{pmatrix} \left[\!\left[\epsilon \right]\!\right] - \frac{1}{\omega^2} \left[\!\left[f_x \right]\!\right] \underline{\alpha} \left[\!\left[f_x \right]\!\right] \underline{\alpha} & -\frac{1}{\omega^2} \left[\!\left[f_x \right]\!\right] \underline{\alpha} \left[\!\left[f_y \right]\!\right] \underline{\beta} \\ -\frac{1}{\omega^2} \left[\!\left[f_y \right]\!\right] \underline{\beta} \left[\!\left[f_x \right]\!\right] \underline{\alpha} & \left[\!\left[\epsilon \right]\!\right] - \frac{1}{\omega^2} \left[\!\left[f_y \right]\!\right] \underline{\beta} \left[\!\left[f_y \right]\!\right] \underline{\beta} \end{pmatrix} \tag{B.25}$$

B.5. smallEigenproblemPML

Implementation of PMLs for isotropic material systems with the structure-transform real-space strategy.

$$\underline{\mathbf{F}} = \begin{pmatrix} \omega^2\,\mu^{11} - \underline{\beta}\,\varepsilon^{33}\,\underline{\beta} & \underline{\beta}\,\varepsilon^{33}\,\underline{\alpha} \\ \underline{\alpha}\,\varepsilon^{33}\,\underline{\beta} & \omega^2\,\mu^{22} - \underline{\alpha}\,\varepsilon^{33}\,\underline{\alpha} \end{pmatrix} \tag{B.26}$$

$$\underline{\mathbf{G}} = \begin{pmatrix} \varepsilon^{22} - \frac{1}{\omega^2}\,\underline{\alpha}\,\mu^{33}\,\underline{\alpha} & -\frac{1}{\omega^2}\,\underline{\alpha}\,\mu^{33}\,\underline{\beta} \\ -\frac{1}{\omega^2}\,\underline{\beta}\,\mu^{33}\,\underline{\alpha} & \varepsilon^{11} - \frac{1}{\omega^2}\,\underline{\beta}\,\mu^{33}\,\underline{\beta} \end{pmatrix} \tag{B.27}$$

with

$$\varepsilon^{11} = \left\lfloor \left[\frac{\epsilon s_y}{s_x} \right] \right\rfloor \tag{B.28a}$$

$$\varepsilon^{22} = \left\lceil \left[\frac{\epsilon s_x}{s_y} \right] \right\rceil \tag{B.28b}$$

$$\varepsilon^{33} = \left[\!\left[\epsilon s_x s_y \right]\!\right]^{-1} \tag{B.28c}$$

and

$$\mu^{11} = \left\lfloor \left[\frac{s_y}{s_x} \right] \right\rfloor \tag{B.29a}$$

$$\mu^{22} = \left\lceil \left[\frac{s_x}{s_y} \right] \right\rceil \tag{B.29b}$$

$$\mu^{33} = \left[\!\left[s_x s_y \right]\!\right]^{-1} \tag{B.29c}$$

Bibliography

[1] E. Yablonovitch. Inhibited Spontaneous Emission in Solid-State Physics and Electronics. *Phys. Rev. Lett.*, **58**: 2059–2062 (May 1987). doi:10.1103/PhysRevLett.58.2059.

[2] S. John. Strong localization of photons in certain disordered dielectric superlattices. *Phys. Rev. Lett.*, **58**: 2486–2489 (Jun 1987). doi:10.1103/PhysRevLett.58.2486.

[3] E. Yablonovitch, T. J. Gmitter, and K. M. Leung. Photonic band structure: The face-centered-cubic case employing nonspherical atoms. *Phys. Rev. Lett.*, **67**: 2295–2298 (Oct 1991). doi:10.1103/PhysRevLett.67.2295.

[4] S. Y. Lin, J. G. Fleming, D. L. Hetherington, B. K. Smith, R. Biswas, K. M. Ho, M. M. Sigalas, W. Zubrzycki, S. R. Kurtz, and J. Bur. A three-dimensional photonic crystal operating at infrared wavelengths. *Nature*, **394**: 251–253 (Jul 1998). doi:10.1038/28343.

[5] S. Noda, N. Yamamoto, H. Kobayashi, M. Okano, and K. Tomoda. Optical properties of three-dimensional photonic crystals based on III–V semiconductors at infrared to near-infrared wavelengths. *Applied Physics Letters*, **75**(7): 905–907 (1999). doi:10.1063/1.124549.

[6] M. Deubel, G. von Freymann, M. Wegener, S. Pereira, K. Busch, and C. M. Soukoulis. Direct laser writing of three-dimensional photonic-crystal templates for telecommunications. *Nature Materials*, **3**: 444–447 (Apr 2004). doi:10.1038/nmat1155.

[7] J. Fischer and M. Wegener. Three-dimensional direct laser writing inspired by stimulated-emission-depletion microscopy. *Opt. Mater. Express*, **1**(4): 614–624 (Aug 2011). doi:10.1364/OME.1.000614.

[8] J. Fischer and M. Wegener. Three-dimensional optical laser lithography beyond the diffraction limit. *Laser & Photonics Reviews* (2012). doi:10.1002/lpor.201100046.

[9] A. Frölich, J. Fischer, T. Zebrowski, K. Busch, and M. Wegener. Complete three-dimensional photonic band gap in the visible. *Nature Materials*. (submitted).

[10] S. G. Johnson and J. D. Joannopoulos. Block-iterative frequency-domain methods for Maxwell's equations in a planewave basis. *Opt. Express*, **8**(3): 173–190 (2001).

[11] M. Deubel, M. Wegener, S. Linden, and G. von Freymann. Angle-resolved transmission spectroscopy of three-dimensional photonic crystals fabricated by direct laser writing. *Applied Physics Letters*, **87**(22): 221104 (2005). doi:10.1063/1.2137899.

[12] P. Russell. Photonic Crystal Fibers. *Science*, **299**(5605): 358–362 (2003). doi:10.1126/science.1079280.

[13] J. C. Knight. Photonic crystal fibres. *Nature*, **424** (Aug 2003). doi:10.1038/nature01940.

[14] K. Busch and S. John. Liquid-crystal photonic-band-gap materials: The tunable electromagnetic vacuum. *Physical Review Letters*, **83**(5): 967–970 (1999).

[15] S. Leonard, J. Mondia, H. Van Driel, O. Toer, S. John, K. Busch, A. Birner, U. Gsele, and

V. Lehmann. Tunable two-dimensional photonic crystals using liquid-crystal infiltration. *Physical Review B - Condensed Matter and Materials Physics*, **61**(4): R2389–R2392 (2000).

[16] C. Schuller, F. Klopf, J. Reithmaier, M. Kamp, and A. Forchel. Tunable photonic crystals fabricated in III-V semiconductor slab waveguides using infiltrated liquid crystals. *Applied Physics Letters*, **82**(17): 2767–2769 (2003).

[17] G. Mertens, T. Rder, H. Matthias, H. Marsmann, H.-S. Kitzerow, S. Schweizer, C. Jamois, R. Wehrspohn, and M. Neubert. Two- and three-dimensional photonic crystals made of macroporous silicon and liquid crystals. *Applied Physics Letters*, **83**(15): 3036–3038 (2003).

[18] D. Noordegraaf, L. Scolari, J. Lægsgaard, L. Rindorf, and T. T. Alkeskjold. Electrically and mechanically induced long period gratings in liquid crystalphotonic bandgap fibers. *Opt. Express*, **15**(13): 7901–7912 (Jun 2007). doi:10.1364/OE.15.007901.

[19] J. D. Jackson. *Classical Electrodynamics* (Wiley, New York, 1999), 3rd edition.

[20] C. Kittel. *Introduction to Solid State Physics* (John Wiley & Sons, Inc., New York, 2005), 8th edition.

[21] N. W. Ashcroft and N. D. Mermin. *Solid State Physics* (Saunders College, Orlando, Fla., 2000), college ed., [repr.] edition.

[22] U. Leonhardt and T. G. Philbin. *Progress in Optics*, volume 53 (Elsevier, 2009), 1st edition.

[23] J. C. Maxwell. A Dynamical Theory of the Electromagnetic Field. *Philosophical Transactions of the Royal Society of London*, **155**: 459–512 (Jan 1865). doi:10.1098/rstl.1865.0008.

[24] D. M. Whittaker and I. S. Culshaw. Scattering-matrix treatment of patterned multilayer photonic structures. *Phys. Rev. B*, **60**: 2610–2618 (Jul 1999). doi:10.1103/PhysRevB.60.2610.

[25] J. H. Poynting. On the Transfer of Energy in the Electromagnetic Field. *Philosophical Transactions of the Royal Society of London*, **175**: 343–361 (1884). doi:10.1098/rstl.1884.0016.

[26] P. C. Clemmow. *The Plane Wave Spectrum Representation of Electromagnetic Fields*, volume 12 of *International Series of Monographs in Electromagnetic Waves* (Pergamon Press, 1966), 1st edition.

[27] E. Hecht. *Optics* (Addison Wesley, 2001), 4th edition.

[28] H. Schade and K. Neemann. *Tensoranalysis* (Walter de Gruyter, Berlin, 2009).

[29] E. J. Post. *Formal Structure of Electromagnetics* (North Holland Publishing Co., Amsterdam, 1962).

[30] T. Ergin, N. Stenger, P. Brenner, J. B. Pendry, and M. Wegener. Three-Dimensional Invisibility Cloak at Optical Wavelengths. *Science*, **328**(5976): 337–339 (Apr 2010). doi:10.1126/science.1186351.

[31] J. Fischer, T. Ergin, and M. Wegener. Three-dimensional polarization-independent visible-frequency carpet invisibility cloak. *Opt. Lett.*, **36**(11): 2059–2061 (Jun 2011). doi:10.1364/OL.36.002059.

[32] International Union of Crystallography: Online Dictionary of CRYSTALLOGRAPHY. Accessed: 30/04/2012.
URL `http://reference.iucr.org/dictionary/`

[33] F. Bloch. Über die Quantenmechanik der Elektronen in Kristallgittern. *Zeitschrift für Physik*

A: Hadrons and Nuclei, **52**: 555–600 (1929). doi:10.1007/BF01339455.

[34] H. Ibach and H. Lüth. *Festkörperphysik: Einführung in die Grundlagen*. Springer-Lehrbuch (Springer, 2008).

[35] I. N. Bronstein, K. A. Semendjajew, G. Musiol, and H. Mühlig. *Taschenbuch der Mathematik* (Verlag Harry Deutsch, 2001), 5th edition.

[36] Wikipedia. Regular Grid. Accessed: 12/09/2012.
URL `http://en.wikipedia.org/wiki/Regular_grid`

[37] CDF Online. Mesh Classification. Accessed: 12/09/2012.
URL `http://www.cfd-online.com/Wiki/Mesh_classification`

[38] Wikipedia. Unstructured Grid. Accessed: 12/09/2012.
URL `http://en.wikipedia.org/wiki/Unstructured_grid`

[39] W. H. Press *et al. Numerical Recipes in C++* (Cambridge University Press, 2002), 2nd edition.

[40] M. Frigo and S. G. Johnson. The Design and Implementation of FFTW3. *Proceedings of the IEEE*, **93**(2): 216–231 (2005). Special issue on "Program Generation, Optimization, and Platform Adaptation".

[41] M. Frigo and S. G. Johnson. *FFTW Documentation Version 3.3.1*. Accessed: May 3rd, 2012.
URL `http://www.fftw.org/fftw3.pdf`

[42] *Intel(R) Math Kernel Library for Linux OS User's Guide*. Accessed: May 3rd, 2012.
URL `http://software.intel.com/sites/products/documentation/hpc/mkl/mkl_userguide_lnx/mkl_userguide_lnx.pdf`

[43] C. Shannon. Communication in the Presence of Noise. *Proceedings of the IRE*, **37**(1): 10–21 (Jan 1949). doi:10.1109/JRPROC.1949.232969.

[44] J. W. Gibbs. Fourier's Series. *Nature*, **59** (Dec 1898). doi:10.1038/059200b0.

[45] J. W. Gibbs. Fourier's Series. *Nature*, **59** (Apr 1899). doi:10.1038/059606a0.

[46] E. Hewitt and R. Hewitt. The Gibbs-Wilbraham phenomenon: An episode in Fourier analysis. *Archive for History of Exact Sciences*, **21**: 129–160 (1979). doi:10.1007/BF00330404.

[47] L. Li. Use of Fourier series in the analysis of discontinuous periodic structures. *J. Opt. Soc. Am. A*, **13**(9): 1870–1876 (Sep 1996). doi:10.1364/JOSAA.13.001870.

[48] G. Bao, L. Cowsar, and W. Masters. *Mathematical Modeling in Optical Science*. Frontiers in Applied Mathematics (Society for Industrial and Applied Mathematics, 2001). Chapter 4.

[49] B. Anic (2011). Private Communication.

[50] L. Li. Reformulation of the Fourier modal method for surface-relief gratings made with anisotropic materials. *Journal of Modern Optics*, **45**(7): 1313–1334 (1998). doi:10.1080/09500349808230632.

[51] L. Li. Fourier modal method for crossed anisotropic gratings with arbitrary permittivity and permeability tensors. *J. Opt. A: Pure Appl. Opt.*, **5**(4): 345–355 (2003). doi:10.1088/1464-4258/5/4/307.

[52] L. Li and G. Granet. Field singularities at lossless metal-dielectric right-angle edges and their ramifications to the numerical modeling of gratings. *J. Opt. Soc. Am. A*, **28**(5): 738–746 (May 2011). doi:10.1364/JOSAA.28.000738.

[53] M. Walz. *B-Spline Modal Method: Eigenmode Solver for Diffractive Optical Systems*. Master's thesis, Karlsruhe Institute of Technology (KIT) (Dec 2011).

[54] B. Anic (Sep 2012). Private Communication.

[55] J. W. S. Lord Rayleigh. On the dynamical theory of gratings. *Proc. R. Soc. Lond. A*, **79**: 399 (1907).

[56] R. W. Wood. Anomalous diffraction gratings. *Phys. Rev.*, **48**: 928–936 (Dec 1935). doi: 10.1103/PhysRev.48.928.

[57] A. W. Snyder and J. D. Love. *Optical Waveguide Theory* (Chapman and Hall Ltd., London, 1983).

[58] V. R. Almeida, Q. Xu, C. A. Barrios, and M. Lipson. Guiding and confining light in void nanostructure. *Opt. Lett.*, **29**(11): 1209–1211 (Jun 2004). doi:10.1364/OL.29.001209.

[59] B. Lutz. *Modeling of point sources in photonic structures via the Fourier modal method*. Master's thesis, Universität Karlsruhe (2009).

[60] P. Yeh, A. Yariv, and E. Marom. Theory of Bragg fiber. *J. Opt. Soc. Am.*, **68**(9): 1196–1201 (Sep 1978). doi:10.1364/JOSA.68.001196.

[61] T. Birks, P. Roberts, P. Russell, D. Atkin, and T. Shepherd. Full 2-D photonic bandgaps in silica/air structures. *Electronics Letters*, **31**(22): 1941–1943 (Oct 1995). doi:10.1049/el: 19951306.

[62] R. F. Cregan, B. J. Mangan, J. C. Knight, T. A. Birks, P. S. J. Russell, P. J. Roberts, and D. C. Allan. Single-Mode Photonic Band Gap Guidance of Light in Air. *Science*, **285**(5433): 1537–1539 (1999). doi:10.1126/science.285.5433.1537.

[63] S. Kawashima, K. Ishizaki, and S. Noda. Light propagation in three-dimensional photonic crystals. *Opt. Express*, **18**(1): 386–392 (Jan 2010). doi:10.1364/OE.18.000386.

[64] I. Staude, G. von Freymann, S. Essig, K. Busch, and M. Wegener. Waveguides in three-dimensional photonic-bandgap materials by direct laser writing and silicon double inversion. *Opt. Lett.*, **36**(1): 67–69 (Jan 2011). doi:10.1364/OL.36.000067.

[65] S. Essig. *Advanced Numerical Methods in Diffractive Optics and Applications to Periodic Photonic Nanostructures*. Ph.D. thesis, Karlsruhe Institute of Technology (KIT) (Feb 2011). URL `http://digbib.ubka.uni-karlsruhe.de/volltexte/1000022240`

[66] A. Yariv, Y. Xu, R. K. Lee, and A. Scherer. Coupled-resonator optical waveguide: a proposal and analysis. *Opt. Lett.*, **24**(11): 711–713 (Jun 1999). doi:10.1364/OL.24.000711.

[67] K. Okamoto. *Fundamentals of Optical Waveguides* (ACADEMIC PRESS, San Diego, USA, 2000).

[68] J. Lundgren. ZEROBESS. Accessed: 31/03/2012. URL `http://www.mathworks.com/matlabcentral/fileexchange/26639`

[69] G. Lecamp, J. P. Hugonin, and P. Lalanne. Theoretical and computational concepts for periodic optical waveguides. *Opt. Express*, **15**(18): 11042–11060 (Sep 2007). doi:10.1364/OE. 15.011042.

[70] P. Monk. *Finite Element Methods for Maxwell's Equations*. Numerical Mathematics and Scientific Computation (Oxford University Press, USA, 2003).

[71] J. C. Nédélec. Mixed finite elements in $\mathbb{R}^3$. *Numerische Mathematik*, **35**: 315–341 (1980). doi:10.1007/BF01396415.

[72] J. C. Nédélec. A new family of mixed finite elements in $\mathbb{R}^3$. *Numerische Mathematik*, **50**: 57–81 (1986). doi:10.1007/BF01389668.

[73] P. Bouchon, F. Pardo, R. Haïdar, and J.-L. Pelouard. Fast modal method for subwavelength gratings based on B-spline formulation. *J. Opt. Soc. Am. A*, **27**(4): 696–702 (Apr 2010). doi:10.1364/JOSAA.27.000696.

[74] M. Walz, T. Zebrowski, J. Küchenmeister, and K. Busch. B-Spline Modal Method: A polynomial approach compared to the Fourier Modal Method. (in preparation).

[75] L. Li. New formulation of the Fourier modal method for crossed surface-relief gratings. *J. Opt. Soc. Am. A*, **14**(10): 2758–2767 (Oct 1997). doi:10.1364/JOSAA.14.002758.

[76] L. Li. Formulation and comparison of two recursive matrix algorithms for modeling layered diffraction gratings. *J. Opt. Soc. Am. A*, **13**(5): 1024–1035 (May 1996). doi:10.1364/JOSAA.13.001024.

[77] T. L. Boullion and P. L. Odell. *Generalized inverse matrices* (Wiley, New York, 1971).

[78] A. Albert. *Regression and the Moore-Penrose Pseudoinverse*. Number 94 in Mathematics in Science and Engineering (Academic Press, New York, 1972).

[79] V. Blobel and E. Lohrmann. *Statistische und numerische Methoden der Datenanalyse*. Teubner Studienbuecher: Physik (Teubner, Stuttgart, 1998).

[80] L. Li. Note on the S-matrix propagation algorithm. *J. Opt. Soc. Am. A*, **20**(4): 655–660 (Apr 2003). doi:10.1364/JOSAA.20.000655.

[81] R. Redheffer. *Difference equations and functional equations in transmission-line theory*. University of California engineering extension series (McGraw-Hill, 1961).

[82] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen. *LAPACK Users' Guide* (Society for Industrial and Applied Mathematics, Philadelphia, PA, 1999), 3rd edition.

[83] R. B. Lehoucq, D. C. Sorensen, and C. Yang. *ARPACK Users' Guide: Solution of Large Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods* (Oct 1997).
URL `http://www.caam.rice.edu/software/ARPACK/`

[84] T. Zebrowski. *Nichtlineare Komplexe Koordinaten Transformationen in der Fourier Moden Methode*. Master's thesis, Universität Karlsruhe (Mar 2008).

[85] C. Klock. *Modeling of point sources in photonic structures via the Fourier modal method*. Master's thesis, Universität Karlsruhe (2009).

[86] T. A. Driscoll and B. Fornberg. A Padé-based algorithm for overcoming the Gibbs phenomenon. *Numerical Algorithms*, **26**: 77–92 (2001). doi:10.1023/A:1016648530648.

[87] T. Driscoll. Singular Fourier-Pade approximation. Accessed: 15/05/2012.
URL `http://www.mathworks.com/matlabcentral/fileexchange/12402`

[88] B. Bai and L. Li. Reduction of computation time for crossed-grating problems: a group-theoretic approach. *J. Opt. Soc. Am. A*, **21**(10): 1886–1894 (Oct 2004). doi:10.1364/JOSAA.21.001886.

[89] P. Mack. *2D H-polarized Auxiliary Basis Functions for the Extension of the Photonic Wannier Function Expansion for Photonic Crystal Circuitry*. Ph.D. thesis, Karlsruhe Institute of Technology (KIT) (Feb 2011).
URL http://digbib.ubka.uni-karlsruhe.de/volltexte/1000023233

[90] M. Lax. *Symmetry Principles in Solid State and Molecular Physics* (Dover Publications, Inc., New York, NY, USA, 1974).

[91] B. Bai and L. Li. Group-theoretic approach to enhancing the Fourier modal method for crossed gratings with one or two reflection symmetries. *Journal of Optics A: Pure and Applied Optics*, **7**(7): 271–278 (2005). doi:10.1088/1464-4258/7/7/002.

[92] B. Bai and L. Li. Group-theoretic approach to the enhancement of the Fourier modal method for crossed gratings: C2 symmetry case. *J. Opt. Soc. Am. A*, **22**(4): 654–661 (Apr 2005). doi:10.1364/JOSAA.22.000654.

[93] B. Benfeng and L. Li. Group-theoretic approach to enhancing the Fourier modal method for crossed gratings of plane group p3. *Journal of Modern Optics*, **52**(11): 1619–1634 (2005). doi:10.1080/09500340500072448.

[94] B. Bai and L. Li. Group-theoretic approach to enhancing the Fourier modal method for crossed gratings with C4 symmetry. *Journal of Optics A: Pure and Applied Optics*, **7**(12): 783 (2005). doi:10.1088/1464-4258/7/12/012.

[95] B. Bai and L. Li. Group-theoretic approach to enhancing the Fourier modal method for crossed gratings with square symmetry. *J. Opt. Soc. Am. A*, **23**(3): 572–580 (Mar 2006). doi:10.1364/JOSAA.23.000572.

[96] L. Li. Field singularities at lossless metal–dielectric arbitrary-angle edges and their ramifications to the numerical modeling of gratings. *J. Opt. Soc. Am. A*, **29**(4): 593–604 (Apr 2012). doi:10.1364/JOSAA.29.000593.

[97] G. Granet. Reformulation of the lamellar grating problem through the concept of adaptive spatial resolution. *J. Opt. Soc. Am. A*, **16**(10): 2510–2516 (Oct 1999). doi:10.1364/JOSAA.16.002510.

[98] T. Vallius and M. Honkanen. Reformulation of the Fourier modal method with adaptive spatial resolution: application to multilevel profiles. *Opt. Express*, **10**(1): 24–34 (Jan 2002).

[99] J. P. Hugonin and P. Lalanne. Perfectly matched layers as nonlinear coordinate transforms: a generalized formalization. *J. Opt. Soc. Am. A*, **22**(9): 1844–1849 (Sep 2005). doi:10.1364/JOSAA.22.001844.

[100] J. Küchenmeister, T. Zebrowski, and K. Busch. A construction guide to analytically generated meshes for the Fourier Modal Method. *Opt. Express*, **20**(16): 17319–17347 (Jul 2012). doi:10.1364/OE.20.017319.

[101] Wikipedia. Jacobian matrix — Inverse. Accessed: 30/05/2012.
URL http://en.wikipedia.org/wiki/Jacobian_matrix_and_determinant#Inverse

[102] Wikipedia. Cramer's Rule. Accessed: 30/05/2012.
URL http://en.wikipedia.org/wiki/Invertible_matrix#Inversion_of_3.C3.973_matrices

[103] T. Weiss, G. Granet, N. A. Gippius, S. G. Tikhodeev, and H. Giessen. Matched coordinates and adaptive spatial resolution in the Fourier modal method. *Opt. Express*, **17**(10): 8051–8061 (May 2009). doi:10.1364/OE.17.008051.

[104] S. Essig and K. Busch. Generation of adaptive coordinates and their use in the Fourier Modal Method. *Opt. Express*, **18**(22): 23258–23274 (Oct 2010). doi:10.1364/OE.18.023258.

[105] A. Taflove and S. Hagness. *Computational Electrodynamics: The Finite-Difference Time-Domain Method*. Artech House Antennas and Propagation Library (Artech House, 2005), 3rd edition.

[106] D. M. Shyroki. Squeezing of Open Boundaries by Maxwell-Consistent Real Coordinate Transformation. *Microwave and Wireless Components Letters, IEEE*, **16**(11): 576–578 (Nov 2006). doi:10.1109/LMWC.2006.884768.

[107] J.-P. Berenger. A perfectly matched layer for the absorption of electromagnetic waves. *Journal of Computational Physics*, **114**(2): 185–200 (1994). doi:10.1006/jcph.1994.1159.

[108] W. C. Chew and W. H. Weedon. A 3D perfectly matched medium from modified maxwell's equations with stretched coordinates. *Microwave and Optical Technology Letters*, **7**(13): 599–604 (1994). doi:10.1002/mop.4650071304.

[109] Z. Sacks, D. Kingsland, R. Lee, and J.-F. Lee. A perfectly matched anisotropic absorber for use as an absorbing boundary condition. *Antennas and Propagation, IEEE Transactions on*, **43**(12): 1460–1463 (Dec 1995). doi:10.1109/8.477075.

[110] W. C. Chew, J. M. Jin, and E. Michielssen. Complex coordinate stretching as a generalized absorbing boundary condition. *Microwave and Optical Technology Letters*, **15**(6): 363–369 (1997). doi:10.1002/(SICI)1098-2760(19970820)15:6⟨363::AID-MOP8⟩3.0.CO;2-C.

[111] F. L. Teixeira and W. C. Chew. Unified analysis of perfectly matched layers using differential forms. *Microwave and Optical Technology Letters*, **20**(2): 124–126 (1999). doi:10.1002/(SICI)1098-2760(19990120)20:2⟨124::AID-MOP12⟩3.0.CO;2-N.

[112] M. Kuzuoglu and R. Mittra. Frequency dependence of the constitutive parameters of causal perfectly matched anisotropic absorbers. *Microwave and Guided Wave Letters, IEEE*, **6**(12): 447–449 (Dec 1996). doi:10.1109/75.544545.

[113] W. von Sellmeier. Zur Erklärung der abnormen Farbenfolge im Spectrum einiger Substanzen. *Annalen der Physik*, **143**: 272–282 (1871).

[114] A. F. Koenderink, A. Lagendijk, and W. L. Vos. Optical extinction due to intrinsic structural variations of photonic crystals. *Phys. Rev. B*, **72**: 153102 (Oct 2005). doi:10.1103/PhysRevB.72.153102.

[115] J. Weirich, J. Lægsgaard, L. Scolari, L. Wei, T. T. Alkeskjold, and A. Bjarklev. Biased liquid crystal infiltrated photonic bandgap fiber. *Opt. Express*, **17**(6): 4442–4453 (Mar 2009). doi:10.1364/OE.17.004442.

[116] J. Weirich, J. Laegsgaard, L. Wei, T. T. Alkeskjold, T. X. Wu, S.-T. Wu, and A. Bjarklev. Liquid crystal parameter analysis for tunable photonic bandgap fiber devices. *Opt. Express*, **18**(5): 4074–4087 (Mar 2010). doi:10.1364/OE.18.004074.

[117] COMSOL, Inc., USA. COMSOL Multiphysics.
URL `http://www.comsol.com`

[118] T. Erdogan. Fiber grating spectra. *Lightwave Technology, Journal of*, **15**(8): 1277–1294 (Aug 1997). doi:10.1109/50.618322.

[119] OpenMPI Team. Open Message Passing Interface.
URL `http://www.open-mpi.org`

Publications

Research Articles in Regular Peer-Reviewed Journals

- *A construction guide to analytically generated meshes for the Fourier Modal Method*; J. Küchenmeister, T. Zebrowski, and K. Busch, Optics Express **20**, pp. 17319–17347 (2012).
- *Direct Transcription of Two-Dimensional Colloidal Crystal Arrays into Three-Dimensional Photonic Crystals*; A. Vlad, A. Frölich, T. Zebrowski, C. A. Dutu, K. Busch, S. Melinte, M. Wegener, and I. Huynen, Advanced Functional Materials (Online). DOI: 10.1002/adfm.201201138
- *Complete three-dimensional photonic bandgap in the visible*; A. Frölich, J. Fischer, T. Zebrowski, K. Busch, and M. Wegener, (Submitted).
- *Simulation of Anisotropic Liquid Crystal Filled Waveguides Using the Fourier Modal Method*; T. Zebrowski and K. Busch, (In Preparation).
- *B-Spline modal method: A polynomial approach compared to the Fourier modal method*; M. Walz, T. Zebrowski, J. Küchenmeister, and K. Busch, (In Preparation).

Conference Proceedings

- *A B-spline modal method in comparison to the Fourier modal method*; M. Walz, T. Zebrowski, J. Küchenmeister, and K. Busch; AIP Conference Proceedings **1398**, pp. 177–179 (2011).

International Conference Oral Presentations

- *Simulation of Liquid Crystal Infiltrated Fibers Using the Fourier Modal Method*; SPIE Photonics Europe, Brussels, (2010)
- *Simulation of Liquid Crystal Based Fiber Devices with the Fourier Modal Method*; 6th Workshop on Numerical Methods for Optical Nano Structures, Zurich, (2010)

International Conference Poster Presentations

- *Photonic Crystal Slab Waveguide Simulations with the Aperiodic Fourier Modal Method*; TaCoNa-Photonics, Bad Honnef, (2008)
- *Simulation of Liquid Crystal Infiltrated Photonic Crystal Fibers Using the Fourier Modal Method*; OSA Optics and Photonic Congress, Karlsruhe, (2010)

Acknowledgements

Naturally, there are many people besides the author who have contributed to this PhD thesis in one way or another over the years. This paragraph is dedicated to those who helped and supported me to successfully finish this thesis.

First off all, my gratitude goes to Prof. Dr. Kurt Busch who gave me the opportunity to work on this thesis and supported me to get the necessary financial ressources. I very much enjoyed the numerous discussions on- and off-topic and the warm atmosphere he created and preseved in his group. Especially in the decisive situations, I could count on his full support and careful advice.

I would like to thank Prof. Dr. Martin Wegener for co-supervising my thesis and for several sucessful collaborations.

It has been a particular joy to work with my colleague Jens Küchenmeister. We had a good time together, especially in the "left over phase" in the end. The numerous fruitful discussions and the close cooperation resulted in great advances in our common project.

My special thanks goes to Andreas Frölich not only for the great cooperation, but also for the common efforts in OSkar with which everything started.

Furthermore, I would like to thank Sabine Essig for introducing me into the world of the FMM and for beeing a great FMM mentor and long-term office mate. She always stayed on top of things, when I already had been lost in the code. I would like to thank as well "my" diploma students Benjamin Lutz and Michael Walz for their great work, mutual help, and a good time together.

For numerous stimulating discussions, I have to thank Michael König and Christian Wolff, with whom the relations extend far beyond work.

For proofreading this thesis, I would like to thank Michael König, Sabine Essig, Christian Wolff, Jens Küchenmeister, Annika Bork, and Aline.

All the small moments that made this long time project endurable can be credited to the entire Photonics Group, which consists of way too many people to mention them all separately. We spend a good time together with daily lunch, barbecues, baggersee and hiking excursions, legendary Funky Photons, poker nights, and so on. Furthermore, they bore long afternoons in test talks to improve them. So, thanks to all!

My six months stay at DTU Copenhagen would not have been possible without my host Jesper Lægsgaard. At DTU, I appreciated much the heartily welcome in the groups of Ole Bang and Karsten Rottwitt. Theose people also introduced me to the fiber optics realm. One person to thank in particular is Johannes Weirich.

I would also like to acknowledge the support of several institutions, namely the Karlsruhe School of Optics & Photonics (KSOP) with their mentoring program, interesting industry excursions, and valuable modules. Thank you to all who took part in this.

Secondly, I am grateful for the ideal and financial support of the Karlsruhe House of Young Scientists (KHYS) which is one of the best benefits of the excellence initiative. In particular I would like to thank Britta Trautwein and Gaby Weick for their help in the application process for the scholarships.

Thirdly and most importantly, I much appreciated the faith in and the financial support of my PhD project by the Carl-Zeiss-Stiftung. Special thanks to Judith Schöffler the heart and soul of the scholarship program.

Last not least, Aline, a thank you would not be enough for beeing by my side through all highs and lows!

Was (sich) lange w(a)ehrt, wird endlich gut.
German proverbial saying.