DISSERTATION
submitted
to the
Combined Faculty for the Natural Sciences and Mathematics
of
Heidelberg University, Germany
for the degree of
Doctor of Natural Sciences

Put forward by

Diplom-Informatiker:	Maximilian Auer
Born in:	Heidelberg
Oral examination:	March 22, 2016

Numerical treatment of localized fields in rigorous diffraction theory and its application to light absorption in structured layers

Dean of the faculty:	Prof. Dr. Michael Gertz,	Heidelberg University
First referee:	Prof. Dr. Karl-Heinz Brenner,	Heidelberg University
Second referee:	Prof. Dr. Wolfgang Osten,	University of Stuttgart
Additional oral examiner:	Prof. Dr. Bernd Jähne,	Heidelberg University

Abstract

This work provides a contribution to the numerical problem of electromagnetic wave diffraction. For this purpose the widely used rigorous coupled-ave analysis (RCWA) is extended to an efficient treatment of incident light beyond the mere plane-wave input. This is essential for modern optical systems with laser or LED sources, which emit finite beams with various profiles. It thereby enables the rigorous analysis of focused light in structured media like in optical storage technology or in optical waveguide coupling. The new extension also dissolves the RCWA's restriction of handling purely linear polarized light sources. Due to a superior truncation scheme the new method inherently conserves energy during the modal propagation even within absorbing grating structures and despite the necessary numerical truncation.
Another part of this work addresses the exact calculation of electromagnetic near fields and local absorption. The results are used to develop a photodetector that is fully compatible with the SOI-CMOS process and does not depend on typical optically active III-V materials. This enables the direct integration of an active, optical component in the manufacturing process of conventional electronic chips and might allow the development of new cost-effective optoelectronic hybrid components in the future, which combine the benefits of both technologies.

Diese Arbeit leistet einen Beitrag zur numerischen Berechnung von Beugungsproblemen. Zu diesem Zweck wird die weit verbreitete Rigorous Coupled Wave Analysis (RCWA) um die Möglichkeit einer effizienten Behandlung von Anregungslicht erweitert, das von der typischerweise angenommenen ebenen Welle abweicht. Dies spielt insbesondere für moderne optische Systeme eine wichtige Rolle, bei denen als Lichtquellen Laser und LEDs zum Einsatz kommen, die Wellenfelder mit endlicher Ausdehnung und unterschiedlichen Strahlprofilen aussenden. Auf diese Weise wird unter anderem die rigorose Behandlung von fokussierten Lichtfeldern in strukturierten Medien ermöglicht, was Anwendungen der RCWA im Bereich der optischer Speichertechnik oder der Wellenleiterkoppelung erlaubt. Durch die neue Erweiterung wird außerdem die Beschränkung der RCWA auf eine rein lineare Polarisation aufgehoben. Indem das Verfahren eine Inkonsistenz in der Trunkierung der numerischen Berechnung auflöst, ist es dem Standardverfahren in Bezug auf die Energieerhaltung in absorbierenden Medien sogar überlegen.
Ein weiterer Teil der Arbeit beschäftigt sich mit der exakten Berechnung von Nahfeldern, sowie der lokalen Absorption. Die Ergebnisse fließen in die Entwicklung eines Photodetektors ein, der vollständig kompatibel zum SOI-CMOS Prozess ist, also auf den Einsatz der typischen optisch aktiven III-V-Materialien verzichtet. Dies ermöglicht die direkte Integration eines aktiven optischen Elementes in den Herstellungsprozess konventioneller elektronischer Chips und könnte somit zur Entwicklung kostengünstiger, optoelektronischer Hybridkomponenten führen, welche die Vorteile beider Technologien vereinen.

Contents

Motivation

Computer simulations have become an essential tool in both science and engineering. They not only provide a time-saving and cost-efficient alternative for many laboratory experiments, but can itself enable scientific discoveries and complex technical developments. In this context, today's simulation techniques assume the function of an experimental setup as well as any measurement instrument in one virtual environment. This allows a significant increase in flexibility, development speed and cost-efficiency. Real experiments are still indispensable, but can be preceded by many virtual iterations. Over the last decades this has changed the way science works, from physics to medicine and even biology. It has also shaped the technological development on all scales from construction via machine engineering through to the rise of the chip technology itself, whose high level of integration would be unthinkable without the constantly evolving performance of computer systems and simulations. Especially with the advances in nanoscience and nanotechnology physical experiments have become significantly more sophisticated and expensive.

This is also true for nanooptics. However, with a deep understanding and the rigorous modeling of the underlying principles of electrodynamics, computer simulation can provide accurate analyses in apparently continuous measurement series. They even allow the visualization of parameters that are difficult and in some cases impossible to measure in a real experiment, like complex field amplitudes or local absorption. In this field, the RCWA has established itself as one of the most important numerical methods to solve the rigorous diffraction problem. The method is known for its high reliability and can be applied to a wide variety of problems. The scope of applications has in fact been successively extended over the years from one-dimensional optical gratings in plane incidence to conical incidence, three-dimensional structures, anisotropic media, non-periodic structures and more. And with every new development the capabilities of computer simulations increase together with the opportunities of nanooptics itself. This has already enabled many developments based on optimized diffractive structures from simple one-dimensional optical gratings[79] used as dispersive elements to complex three-dimensional photonic crystals for sophisticated wave guidance in next-generation signal-processing devices[54].

A typical, although not fundamental, restriction of the RCWA, so far, is the consideration of pure plane wave illumination for the diffraction problem. However, real light sources like lasers and LEDs usually emit finite beams with various profiles. Most optical setups make use of focused light and even plane waves are often distorted in real experiments. So, there is an apparent potential for another improvement of the method, which is the extension to arbitrarily shaped input fields. An efficient implementation of this idea is developed in chapter 3.

Nanooptics is also a rapidly growing market, but despite its great potential it still struggles to establish itself in the consumer sector unlike modern electronics, which appear omnipresent. One important reason is the difficulty to integrate optical technology into the manufacturing process of standard electronics. A solution to this problem would significantly reduce development and fabrication costs, since a large part of the existing infrastructure could be reused, and would accelerate the advent of optical nanotechnology. Chapter 4 and 5 investigate techniques to calculate the exact near fields and local absorption to finally develop a photodetector as an active optical element that is fully compatible to the SOI-CMOS process in chapter 6.

Chapter 1

Fundamentals

Maxwell's equations together with the constitutive equations and the Lorentz force law form the basis of classical, non-quantum, electrodynamics. They describe the interplay between electric and magnetic waves and their interaction with charges and currents. They also provide the mathematical foundation for phenomena of higher abstraction like wave propagation and the interaction with structured matter. In fact, these fundamental principles can be used not only to understand, but also to predict all classical electrodynamic phenomena, ranging from free propagation, reflection and absorption to refraction and diffraction.

However, the descriptive character of these equations is very fundamental. Their solution for a specific situation is the purpose of diffraction theory and also of the rigorous coupled wave analysis (RCWA), which is one of several numerical solution approaches and who's closer study (cf. Chap. 2) and extension (cf. Chap. 3) is an important part of this work. The present chapter summarizes the basics of electrodynamics, which are necessary for the understanding of this work, and introduces some concepts of diffraction theory, which are used by the RCWA and related methods. The presented derivations are based on the descriptions by Brenner[14], Born and Wolf[11] and further references, which are provided in the text. This chapter also provides the basis for the calculation of local absorption (cf. Chap. 5), which is used in the review of the RCWA's conservation of energy and the design and optimization of a silicon based photodetector (cf. Chap. 6). At the end of this chapter some alternative approaches for the diffraction problem are given in a brief review.

1.1 Electrodynamic principles

1.1.1 Maxwell's equations

Initially in 1961, Maxwell[81] published a set of twenty equations, which summarized the knowledge about electricity and magnetism at that time. He was thereby able to prove the interconnection between both domains and proposed that light was combination of electric and magnetic waves influencing each other. Later, Heaviside[46] and concurrently Gibbs[39] and Hertz[49, 50] condensed the equations into today's form of only four using vector notation.

Their differential form is

1. Ampère's circuital law	$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial}{\partial t}\mathbf{D}$	(1.1a)
2. Maxwell–Faraday equation	$\nabla \times \mathbf{E} = -\frac{\partial}{\partial t}\mathbf{B}$	(1.1b)
3. Gauss's law	$\nabla \cdot \mathbf{D} = \rho$	(1.1c)
4. Gauss's law for magnetism	$\nabla \cdot \mathbf{B} = 0$	(1.1d)

All of these quantities are functions of position $\mathbf{r} = (x\ y\ z)^T$ in space and time t.

The Ampère's circuital law states that a magnetic field $\mathbf{H}(\mathbf{r}, t)$ is either generated by an electric current or by the change of an electric field. In this relationship $\mathbf{J}$ denotes the electric current density and $\mathbf{D}(\mathbf{r}, t)$ is the electric displacement field[1]. According to the Maxwell–Faraday's law an electric field $\mathbf{E}(\mathbf{r}, t)$ is induced by a time varying magnetic field (magnetic induction $\mathbf{B}(\mathbf{r}, t)$). The Gauss's law describes electric charges to be the source of static electric fields. Here, ρ is a charge density. Finally, the Gauss's law for magnetism states that magnetic fields are source-free, or in other words, there are no magnetic charges, but dipoles only.

One important property of all Maxwell's equations is their linearity. Due to the superposition principle any linear combination of a given solution will lead to another valid solution for these equations.

1.1.2 Material properties

While electric and magnetic field amplitudes ($\mathbf{E}$ and $\mathbf{H}$) are describing external fields, electric displacement and magnetic induction ($\mathbf{D}$ and $\mathbf{B}$) describe the interplay between an external field with polarisable and magnetizable matter. The constitutive equations express the relationship between these quantities:

$$\mathbf{D} = \epsilon_0 \epsilon \mathbf{E} \tag{1.2a}$$

$$\mathbf{B} = \mu_0 \mu \mathbf{H} \tag{1.2b}$$

ϵ_0 and μ_0 are called electric and magnetic constants or the permittivity and the permeability of the vacuum. ϵ and μ are their correspondent material properties in any other arbitrary medium. They describe how much the creation of an electric or magnetic field is affected by the medium.

[1] The electric displacement field $\mathbf{D}$, also known as Maxwell's displacement current, was not part of the original formulation of the Ampère's circuital law. It was first added by Maxwell to obtain a form of the law that is consistent with the continuity equation.

For linearly responding materials the constitutive equations can also be written as

$$\mathbf{D} = \epsilon_0 (1 + \chi_e) \mathbf{E} = \epsilon_0 \mathbf{E} + \mathbf{P} \tag{1.3a}$$
$$\mathbf{B} = \mu_0 (1 + \chi_m) \mathbf{H} = \mu_0 (\mathbf{H} + \mathbf{M}), \tag{1.3b}$$

where the electric susceptibility χ_e indicates the degree of polarization to an electric field and the magnetic susceptibility χ_m the magnetization to a magnetic field. The aggregated quantities on the right hand side are called polarization density $\mathbf{P}$ and magnetization $\mathbf{M}$. The frequency dependency of these quantities is called dispersion.
On the subatomic level, an external electromagnetic field separates negative electrons from positive atom nuclei, inducing an electric dipole moment $\mathbf{p}$. N dipoles in a unit volume then form the electric polarization $\mathbf{P} = N\mathbf{p}$. However, polarization is not an instantaneous effect, but obeys causality. Thus, the atomic and molecular processes, which are involved in the medium's response to an external electric field, cause damping forces, which lead to a delay between the electric field $\mathbf{E}$ and the polarization $\mathbf{P}$ and also to a loss of electromagnetic energy as absorption. Consequently dispersion is closely related to absorption as described by the Kramers-Kronig relation[55].
On a macroscopic scale the phase-shift between $\mathbf{P}$ and $\mathbf{E}$ and therefore between $\mathbf{D}$ and $\mathbf{E}$ can be expressed by a complex permittivity ϵ, where the imaginary part ϵ_i represents the absorptive property of the material:

$$\epsilon = \epsilon_r + i\epsilon_i \tag{1.4}$$

Generally, corresponding processes in magnetic and conducting media can also cause phase shifts between $\mathbf{E}$ and $\mathbf{J}$ or $\mathbf{B}$ and $\mathbf{H}$, leading to a complex permeability μ.

Another material property is the conductivity σ. In all considered cases the external electric field and the current density obey a linear relationship that can be expressed by the Ohm's law

$$\mathbf{J} = \sigma \mathbf{E}. \tag{1.5}$$

However, in Sec. 1.1.3 it is shown, that the electric current density can usually be neglected for optical frequencies and thus all materials can be fully described by the complex and position dependent quantities permittivity $\epsilon(x, y, z)$ and permeabilities $\mu(x, y, z)$.

Remark: In anisotropic media the material properties ϵ, μ and σ are potentially direction dependent, which can be expressed by a tensor notation. However, this work concentrates on isotropic media only and hence a scalar description is sufficient.

1.1.3 Harmonic fields

Monochromatic fields with a sinusoidally time dependency are called harmonic. They arise from harmoniously oscillating sources and form in linear media once they have reached a steady state. Mathematically this temporal dependency can be described by a single angular frequency ω using the complex exponential notation based on Euler's formula:

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}(\mathbf{r})\, e^{-i\omega t}. \tag{1.6}$$

Throughout this work electromagnetic fields are all considered monochromatic and harmonic, assuming a steady state in a linear medium. However, by applying the inverse Fourier transform a general solution of Maxwell's equations can be built by a (coherent) superposition of multiple single-frequency solutions using the inverse Fourier transform:

$$\mathbf{E}(\mathbf{r}, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{\mathbf{E}}(\mathbf{r}, \omega)\, e^{-i\omega t} d\omega \tag{1.7}$$

Harmonic fields considerably simplify Maxwell's equation since the time derivative $\partial/\partial t$ can simply be replaced by $-i\omega$, leading to

$$\nabla \times \mathbf{H} = \sigma \mathbf{E} - i\omega \mathbf{D} \tag{1.8a}$$
$$\nabla \times \mathbf{E} = i\omega \mathbf{B} \tag{1.8b}$$
$$\nabla \cdot \mathbf{D} = \rho \tag{1.8c}$$
$$\nabla \cdot \mathbf{B} = 0. \tag{1.8d}$$

Thus, without loss of generality a fixed time like $t = 0$ can be chosen to remove any time dependencies for all following considerations.

1.1.4 Wave equations

The combination of the first two Maxwell equations 1.8a and 1.8b and a substitution of the **D**- and **B**-fields according to the constitutive equations 1.2a and 1.2b results in the inhomogeneous wave equations. For harmonic electric and magnetic fields in charge-free media ($\rho = 0$) they can be written as

$$\nabla^2\mathbf{E} + \mu\mu_0\epsilon_0\omega^2\left(\epsilon + i\frac{\sigma}{\epsilon_0\omega}\right)\mathbf{E} + \nabla\left(\ln\left(\mu\right)\right) \times \left(\nabla \times \mathbf{E}\right) + \nabla\left(\mathbf{E}\nabla\ln\left(\epsilon\right)\right) = 0 \tag{1.9a}$$

and

$$\nabla^2\mathbf{H} + \epsilon\epsilon_0\mu\mu_0\omega^2\mathbf{H} + \nabla\left(\ln\left(\epsilon\right)\right) \times \left(\nabla \times \mathbf{H}\right) + \nabla\left(\mathbf{H}\nabla\ln\left(\mu\right)\right) = 0. \tag{1.9b}$$

These two differential equations represent a general formulation for the propagation of light in any linear and charge-free medium.
With Maxwell's addition the Ampère's circuital law (cf. Eqs. 1.1a and 1.8a) predicts two sources of the magnetic field: the electric current and the displacement current. Since the displacement current density $\partial\mathbf{D}/\partial t = \epsilon\epsilon_0\partial\mathbf{E}/\partial t$ is proportional to the time derivative of the electric field, they are always out of phase. This can again be expressed by a complex permittivity $\epsilon + i\frac{\sigma}{\epsilon_0\omega}$. However, since even for metals the conductivity σ is only in the order of magnitude of 10^6, its influence can usually be neglected for optical frequencies.
Inhomogeneities in the medium manifest in the gradient terms of Eqs. 1.9a and 1.9b and cause a coupling of the different field components. This makes an analytical solution difficult and practically impossible in the general case and gave rise to a large number of numerical approaches. A further discussion on this topic is given in Secs. 1.2.2 and 1.2.3.
Assuming constant material properties ϵ and μ will let the gradient terms in Eqs. 1.9a and 1.9b disappear. Together with the following associations

$$\text{Speed of light in vacuum:} \quad c_0 = \frac{1}{\sqrt{\epsilon_0\mu_0}} \tag{1.10a}$$

$$\text{Refractive index (Maxwell relation):} \quad n = \sqrt{\epsilon\mu} \tag{1.10b}$$

$$\Rightarrow \text{Speed of light in a medium:} \quad c = \frac{1}{\sqrt{\epsilon_0\mu_0\epsilon\mu}} \tag{1.10c}$$

the homogeneous form of the harmonic wave equation for the electric field can be derived as

$$\nabla^2\mathbf{E} + \frac{\omega^2}{c^2}\mathbf{E} = 0, \tag{1.11}$$

where different the field components appear uncoupled. A corresponding equation can equally be deduced for the magnetic field.

1.1.5 Plane waves

The concept of a plane wave describes a three-dimensional harmonic wave with a single spatial frequency ν and wavefronts, which are indefinitely extended perpendicular to a single direction of propagation. It is defined as

$$\mathbf{E}(\mathbf{r}, t) = E_0 \mathbf{E_1} e^{i(\mathbf{kr} - \omega t)}, \tag{1.12}$$

with E_0 as a constant field amplitude and $\mathbf{E_1}$ as the normalized polarization vector. $\mathbf{k} = (k_x \; k_y \; k_z)^T$ is called the wave vector and defines the direction of propagation. It is related to the frequency ν and the wavelength λ through the wave number $k = |\mathbf{k}| = 2\pi\nu = 2\pi/\lambda$. It can easily be verified, that the plane wave definition 1.12 is also a solution of the wave equation (cf. Eq. 1.11), but only if the dispersion relation

$$\text{Wave number in vacuum:} \qquad k_0 = \frac{\omega}{c_0} \tag{1.13a}$$

$$\text{Wave number in a medium:} \quad k = nk_0 = n\frac{\omega}{c_0} = \frac{\omega}{c} \tag{1.13b}$$

holds. Thus, the wave vector can also be expressed as

$$\mathbf{k} = nk_0\mathbf{s} \tag{1.14}$$

with $\mathbf{s}$ defined in spherical coordinates (cf. App. A.1) as the unit vector of $\mathbf{k}$.

Finally, the dispersion relation can also be used to derive the Helmholtz equation, which is the scalar correspondent to the homogeneous wave equation (cf. Eq. 1.11)

$$\nabla^2 E_0 + k^2 E_0 = 0 \tag{1.15}$$

However, such a scalar description is only valid in those cases where the coupling between the electric and magnetic field vector components can be neglected like in a homogeneous medium.

1.1.6 Angular spectrum method

Since plane waves satisfy the wave equation, so does any linear combination of them. This fact can be used to propagate any electric field in a homogeneous medium. The idea is to decompose the field into plane waves using Fourier transform, then propagate the plane waves individually and finally superpose the results through inverse Fourier transform to obtain the propagated electric field.

The definition of an ideal plane wave describes the electric field at any position of space. Substituting the z-component of the wave vector $\mathbf{k}$ as

$$k_z = \pm\sqrt{k^2 - k_x^2 - k_y^2} \tag{1.16}$$

with $k = \omega/c$ according to Eq. 1.13b, the definition of the plane wave can be rewritten as

$$\begin{aligned}
\mathbf{E}(\mathbf{r}_\perp, z) &= E_0\mathbf{E_1}\; e^{i\mathbf{k}_\perp\mathbf{r}_\perp} e^{\pm iz\sqrt{k^2-k_x^2-k_y^2}} && \text{(1.17a)}\\
&= \mathbf{E}(\mathbf{r}_\perp, 0)\; e^{\pm iz\sqrt{k^2-k_\perp^2}} && \text{(1.17b)}\\
&= \mathbf{E}(\mathbf{r}_\perp, 0)\; P^{\pm}(k_\perp, z). && \text{(1.17c)}
\end{aligned}$$

Eq. 1.17c describes the propagation of plane wave from the *plane of origin* (the x/y-plane at $z = 0$) along the z-axis using the definitions $\mathbf{r}_\perp = (x\; y)^T$ and $\mathbf{k}_\perp = (k_x\; k_y)^T$.
$P^{\pm}(k_\perp, z) = \exp\left(\pm iz\sqrt{k^2 - k_\perp^2}\right)$ can be identified as a propagator term. A positive sign of z indicates wave propagation in forward direction (positive z-direction) and a negative sign stands for propagation in backward direction. However, position and orientation of the coordinate system is arbitrary and without loss of generality only the forward propagation is considered here.
Nonetheless, for $k_z = \sqrt{k^2 - k_\perp^2}$ three cases need to be distinguished:

$$\begin{aligned}
k_\perp^2 \leq (nk_0)^2 \quad &\Rightarrow k_z \in \Re \quad \Rightarrow \text{propagating planar wave} && \text{(1.18a)}\\
k_\perp^2 = (nk_0)^2 \quad &\Rightarrow k_z = 0 \quad \Rightarrow \text{cut-off frequency} && \text{(1.18b)}\\
k_\perp^2 > (nk_0)^2 \quad &\Rightarrow k_z \in \Im \quad \Rightarrow \text{evanencent wave} && \text{(1.18c)}
\end{aligned}$$

Equation 1.18a describes spatial frequencies with only a small lateral component. They lead to propagative waves in the far field. Frequencies beyond the cut-off frequency, on the other hand, can only propagate in the plane at $z = 0$ (cf. Eq. 1.18c). Along the z-axis, they are exponentially damped and do not carry energy.

The angular spectrum method now uses the two-dimensional Fourier transform

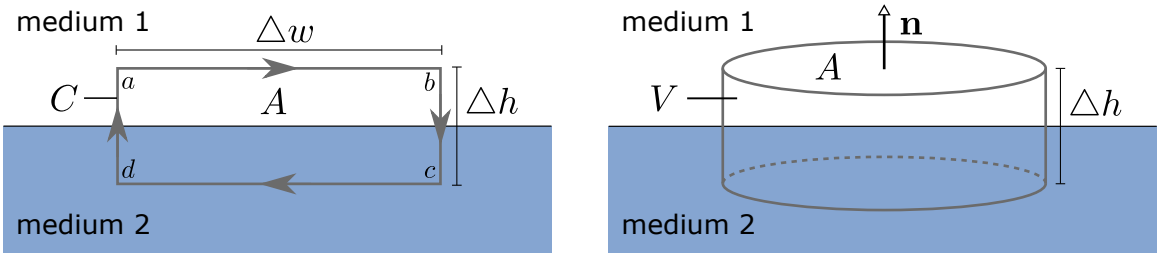
$$\tilde{\mathbf{E}}(\mathbf{k}_\perp, 0) = \iint \mathbf{E}(\mathbf{r}_\perp, 0)\, e^{-i\mathbf{k}_\perp \mathbf{r}_\perp} d^2\mathbf{r}_\perp \tag{1.19}$$

to decompose the electric field in the plane of origin into plane waves and appends the plane wave propagator (cf. Eq. 1.17c) to the inverse Fourier transform:

$$\mathbf{E}(\mathbf{r}_\perp, z) = \iint \tilde{\mathbf{E}}(\mathbf{k}_\perp, 0)\, \underbrace{e^{iz\sqrt{k^2-k_\perp^2}}}_{P^+(k_\perp, z)}\, e^{i\mathbf{k}_\perp \mathbf{r}_\perp} \frac{d^2 k_\perp}{(2\pi)^2}. \tag{1.20}$$

This way, any field distribution, which is known in the plane of origin, can be propagated by the principles of plane wave decomposition and superposition, while the propagation is performed on the individual plane waves in Fourier space. With $\mathbf{k}_\perp = 2\pi\nu_\perp$ the tangential wave vector $\mathbf{k}_\perp$ can be identified as a coordinate in the angular spectrum. This fact also gave rise to the name angular spectrum method (ASM).

1.1.7 Maxwell's boundary conditions



(a) Closed path C through two homogeneous media

(b) A volume V separated by a material interface

Figure 1.1: Illustration of the integration paths of Maxwell's equations across a material interface

Maxwell's equations can also be expressed in an integral form. Its equivalence to the differential form is given by the Kelvin–Stokes theorem for the Ampère's circuital law (cf. Eq. 1.1a) and the Maxwell–Faraday equation (cf. Eq. 1.1b) and given by the divergence theorem for the Gauss's law (cf. Eq. 1.1c) and the Gauss's law for magnetism (cf. Eq. 1.1d).

The first two Maxwell equations describe the relation between a path integral over a path C and a surface integral over a surface A:

$$\oint_C \mathbf{H}\, d\mathbf{l} = \iint_A \mathbf{j}\, d\mathbf{A} + \frac{\partial}{\partial t} \iint_A \mathbf{D}\, d\mathbf{A} \tag{1.21a}$$

$$\oint_C \mathbf{E}\, d\mathbf{l} = -\frac{\partial}{\partial t} \iint_A \mathbf{B}\, d\mathbf{A}, \tag{1.21b}$$

if the path C is considered in a way that it encloses the boundary between two homogeneous media, as illustrated in Fig. 1.1a, then in the limit as $\triangle h$ tends to zero the area A disappears and for finite arguments the surface integrals on the right sides of Eqs. 1.21a and 1.21b tend to zero. On the left side, however, only the normal components disappear while the tangential components $\mathbf{F}_\perp$ with $\mathbf{F} \in \{\mathbf{E}, \mathbf{H}\}$ remain:

$$\oint_C \mathbf{F}_\perp\, d\mathbf{l} = 0 \tag{1.22a}$$

$$\int_a^b \mathbf{F_1}_{;\perp} d\mathbf{l} - \int_c^d \mathbf{F_2}_{;\perp} d\mathbf{l} = 0 \tag{1.22b}$$

$$\mathbf{F_1}_{;\perp} \triangle w - \mathbf{F_2}_{;\perp} \triangle w = 0 \tag{1.22c}$$

$$\Rightarrow \mathbf{F_1}_{;\perp} = \mathbf{F_2}_{;\perp} \tag{1.22d}$$

This leads to the conclusion that

The tangential components of the electric and the magnetic field* $\mathbf{E}$ *and* $\mathbf{H}$ *are continuous. [1]

The integral form of Maxwell's third and fourth equation reads as:

$$\oiint_S \mathbf{D}\, d\mathbf{A} = \iiint_V \rho\, dV \tag{1.23a}$$

$$\oiint_S \mathbf{B}\, d\mathbf{A} = 0. \tag{1.23b}$$

Here, S is a closed surface of a volume V, which in case of Eq. 1.23a encloses an electric charge. Considering a similar case like before, in which the integration volume encloses the material interface as depicted in Fig. 1.1b, another continuity condition can be derived: In the limit as $\triangle h \to 0$ the volume V tends to zero, the electric charge disappears on the right side of Eq. 1.23a and the electric and magnetic flux can only pass through the remaining surfaces $d\mathbf{A} = \pm A \cdot \mathbf{n}$. Hence, in analogy to the previous case, one can conclude:

The normal components of the electric displacement field and magnetic induction* $\mathbf{D}$ *and* $\mathbf{B}$ *are continuous. [2]

1.2 Diffraction theory

Diffraction describes the phenomenon of non-geometric propagation of light in the encounter of structures of a size similar to the wavelength. These structures can be edges of apertures, but also complex near-wavelength geometries as well as media with inhomogeneous material properties. The interaction causes disturbances in the wavefronts, which propagate equally in all radial directions (as described by the Huygens-Fresnel principle) and result in interference patterns and light propagation in geometric shadow areas.
However, the influence of diffraction effects decreases with a growing size of the structure, since the diffractive features of the object represent a smaller proportion of the overall structure. Thus, for macroscopic structures, the effects can usually be neglected and the interaction between light and objects can be described by geometrical optics.
For objects with features sizes of only few wavelengths light shows its electromagnetic wave character and diffraction effects become significant. This requires the solution of the Maxwell equations in every position of space including some finite or infinite boundaries. However, for many applications the description can be limited to the propagation of the field in a homogeneous medium outside the diffractive structure. Since the wave vector components do not couple there, the vector character of the wave equation can be neglected and a scalar description of the fields is sufficient. In this case the interaction with the diffractive structure can be reduced to the direct influence on the phase and amplitude of the light outside the structure while ignoring the complex behavior in the inside.
If the size of the diffractive structures approaches the size of the wavelength, Maxwell's equations need to be solved in a rigorous way with respect to its full vector nature. At material boundaries or in inhomogeneous media the gradients in the material properties lead to a coupling of the vector field components and directly influence the field distribution in the near surrounding - the so called near field. However, the interaction on this scale can also influence the far field in larger distance to the object, especially due to resonant effects and periodic structures. Although this domain requires the most complex methods, it also enables many interesting and valuable applications (cf. Sec. 1.2.1). Since evanescent waves (with $P(k_\perp, z) = \exp(ik_z z)$, $k_z \in \Im$, cf. Eq. 1.18c) decay exponentially with distance they are primarily a near field phenomenon, and due to the periodicity of most diffractive structures in technical applications, also the far field can in fact be a well-structured. This even allow predictions about the far field that do not even rely on a solution of the wave equation. One example is the grating equation (cf. Sec. 1.2.1) that can be used in geometric optics as a replacement of Snell's law for refraction.
Finally, structures with feature sizes significantly smaller than the wavelength do not show diffraction effects anymore as they appear invisible to the light. They can be approximated with effective medium approximations (EMA)[2], which assumes an averaged refractive index of the sub-lambda structure.

1.2.1 Diffraction at Gratings

Optical gratings are a systematic way to utilize diffraction effects for various applications. They are characterized by a periodic structure or medium and thereby also give structure to the far field. This is especially, but not only, true for collimated and coherent incident light. For applications the most important diffraction phenomenon is the appearance of discrete diffraction orders, which were first discovered by Rittenhause[108] in 1786 and further elaborated by Young[137]. Another important fact is that the deflection angle of the diffracted light outside the specular order (zero-order mode) is wavelength depended.
As outlined by Popov[106, p. 7], this makes a diffraction grating a dispersive optical component with several advantages over a prism: First, *the grating can be a plane device, while the prism is a bulk one that requires larger volumes of optically pure glass (to add the difficulties of weight and temperature expansion constraints.* Secondly, *provided a suitable reflecting material, the grating can work in spectral regions, where there is no transparent 'glass' with sufficient dispersion.* And thirdly, *grating dispersion can be varied, as it depends on the groove period, while prism dispersion depends on the material choice and groove angle, which gives quite limited choices.*

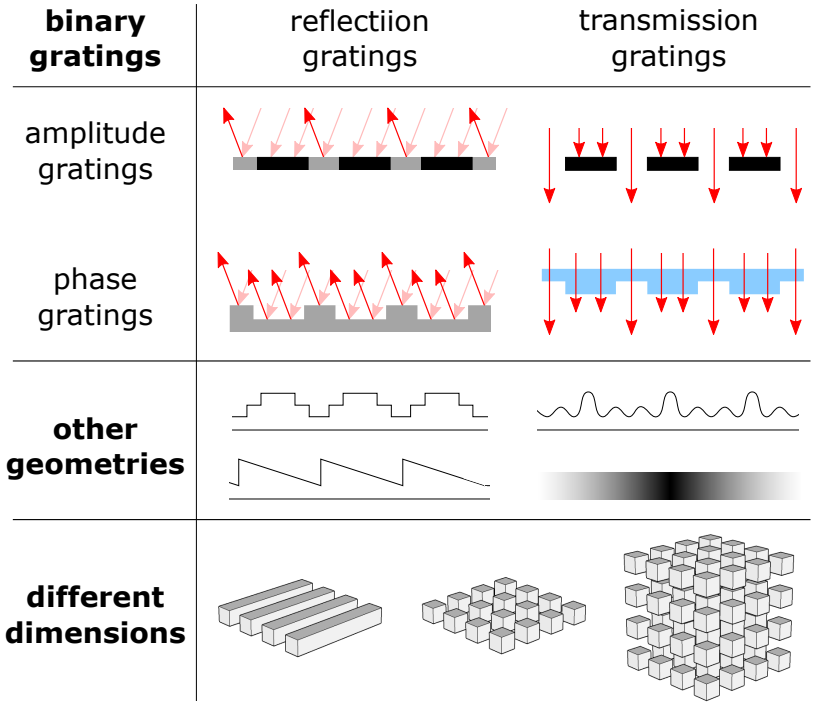


Figure 1.2: Different types of diffraction gratings

There are various types of diffraction grating: A distinction is made between reflective and transmissive diffraction gratings. There are gratings that influence phase and amplitude exclusively or jointly. They also come in various geometries, which range from binary gratings over multi-level gratings, triangular shaped - so called blazed - grating to continuous surface profiles and even gradient index gratings. And finally diffraction gratings can be one-, two- and three-dimensional. Figure 1.2 shows an overview.

Applications

The number of applications for diffraction gratings is in fact vast. It includes spectral analyses[79, Chap. 1], beam splitting [70], simple wavelength blocking filters (cf. Sec. 6.4) or guided-mode resonance filters[130], coupling in integrated optics[122], subwavelength gratings for enhanced local absorption in solar cells[7, 47] or photodetectors (cf. Secs. 6.2 and 6.3), excitation of surface plasmon polaritons[51] and not to mention the potential of photonic crystals as the optical counterpart to semiconductor crystals for electrons in terms of band gap structures[99, 129]. A good overview is given by Loewen and Popov[79, 106].

Polarization

Polarization of an electromagnetic wave is defined as the orientation of its electric and magnetic field vector. In free space electromagnetic waves are transversal, meaning that both, the electric field vector $\mathbf{E}$ and the magnetic field vector $\mathbf{H}$ oscillate linearly or circularly in a plane perpendicular to the propagation vector $\mathbf{k}$. For plane waves, this can easily be shown by substituting the definition of the plane wave (cf. Eq. 1.12) into the first or second Maxwell's equations (cf. Eqs. 1.8a and 1.8b).

Example:

$$\mathbf{k} \times \mathbf{E} = \omega\mu\mu_0\mathbf{H} \quad \Rightarrow \quad \mathbf{k} \perp \mathbf{E} \perp \mathbf{H} \tag{1.24}$$

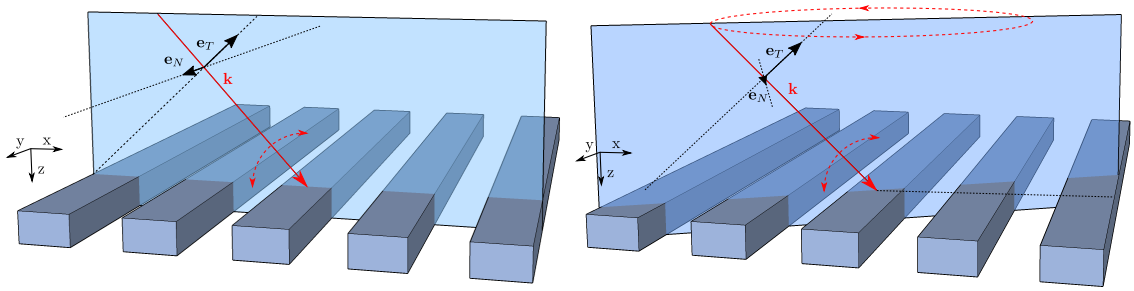
This implies that the different vector components are not independent and each one can be expressed by the other, e.g.:

$$E_z = -\frac{k_x E_x + k_y E_y}{k_z} \quad \text{for } k_z \neq 0. \tag{1.25}$$

Usually polarization becomes relevant not before any kind of interaction with an anisotropic medium or with the surface of a geometric object. If in the latter case the object is a one-dimensional diffraction grating and the incident light lies in a plane perpendicular to the grating grooves - a configuration known as classical mounting, then it is convenient to distinguish between two types of polarization: At TE polarization, the electric field vector is tangential to the grating grooves, while at TM polarization, the magnetic field vector is (cf. Fig. 1.3a). With the typical convention, that the grating grooves are aligned to the x-axis and the grating's normal vector is pointing in the direction of z, the partial derivations with respect to y disappear in Maxwell's equations. This leads to two disjunct sets of equations, which can be solved separately for each type of polarization[97, Chap. 1.2.2].

A general polarization vector $\mathbf{E_1}$ of the incident plane wave $\mathbf{E} = E_0\mathbf{E_1}e^{i\mathbf{kr}}$, can be defined by an angle ψ in a local coordinate system that is spanned by two unit vectors $\mathbf{e}_T$ and $\mathbf{e}_N$, tangential and normal to the plane of incidence:

$$\mathbf{E_1} := sin(\psi)\mathbf{e}_N + cos(\psi)\mathbf{e}_T. \tag{1.26}$$



(a) Classical mounting (b) Conical mounting

Figure 1.3: Figures (a) and (b) show the polarization basis vectors for classical and conical mounting. In classical mounting the electromagnetic field is called TE polarized if the electric field vector $\mathbf{E}$ is aligned with $\mathbf{e}_N$ (or $\phi = 0$ deg). If one the other hand the electric field vector is aligned with $\mathbf{e}_T$ (or $\phi = 90$ deg), which leads to an alignment between the magnetic field and $\mathbf{e}_N$, then the field is called TM polarized.

However, if either the grating becomes two-dimensional (with an additional y-dependency of the material properties) or the incident light does not lie in the described plane perpendicular to the grating grooves, then the distinction of the two types of polarization is no longer justified. If the the plane of incidence is rotated around the z-axis and the grating is one-dimensional (cf. Fig. 1.3b), the diffraction orders are aligned on the surface of a cone and thus the setup is referred to as conical mounting.
Nevertheless, relation 1.26 is still applicable in the general case, if the two polarization base vectors $\mathbf{e}_N$ and $\mathbf{e}_T$ are also defined for non-classical mounting. Such definition can easily be derived from the assumptions $\mathbf{k} \perp \mathbf{e}_z \perp \mathbf{e}_N$ and $\mathbf{k} \perp \mathbf{e}_N \perp \mathbf{e}_T$ respectively:

$$\mathbf{e}_N = \frac{\mathbf{e}_z \times \mathbf{s}}{|\mathbf{e}_z \times \mathbf{s}|} = \frac{1}{s_\perp}\begin{pmatrix} -s_y \\ s_x \\ 0 \end{pmatrix} \tag{1.27a}$$

$$\mathbf{e}_T = \frac{\mathbf{e}_N \times \mathbf{s}}{|\mathbf{e}_N \times \mathbf{s}|} = \frac{1}{s_\perp}\begin{pmatrix} s_x s_z \\ s_y s_z \\ -s_\perp^2 \end{pmatrix}. \tag{1.27b}$$

Again, $\mathbf{s}$ is the unit vector of the wavevector $\mathbf{k}$. For normal incidence, where $s_\perp = 0$ the definition must be set to $\mathbf{e}_N = \mathbf{e}_y$ and $\mathbf{e}_T = \mathbf{e}_x$ to be compatible with classical mounting.
Besides the here described forms of linear polarization, more complex forms become available, when deviating from the simple plane wave illumination. Those are discussed in Sec. 3.9.2.

Diffracted fields of optical gratings

The periodic structure of the grating concept also helps to describe diffraction in a mathematical sense. The following section gives a short formal justification for the intuitive fact, that the diffracted field of an optical grating should reflect the grating's periodic nature. It basically describes how so called pseudo-periodic fields arise from the periodic boundary conditions in case of a tilted illumination. This allows the separation of space variables using a special form of a Fourier series as predicted by the Floquet theorem. The following derivation can also described in similar form by Petit[103] and Neviere[97, Chap. 1.2.3-5], Trauter[126] and Hench et al.[48] on which the following description is based on.

In the considered case the interface between two media is described by a one-dimensional periodic function $f(x) = f(x + P_x)$. The light falls through a transparent medium I onto the surface of an ideally conducting medium II. The result is a perfect reflective grating with no electromagnetic field in medium II. This example can easily be extended to more general cases[103]. The incident plane wave is TE polarized in the classical sense (cf. Sec. 1.2.1) with a wave vector $\mathbf{k}_i$ laying in the x/z-plane ($\Rightarrow k_{i;y} = 0$). Thus the electric field only consists of a scalar y-component and can be written as

$$E_i(x,z) = E_0 \underbrace{e^{ik_{i;z}z}}_{P(z)} \underbrace{e^{ik_{i;x}x}}_{e^{i\phi(x)}} \tag{1.28}$$

with a constant amplitude E_0, a propagator $P(z)$ and a phase term $\exp(i\phi(x))$ to account for a tilt of the incident wave[2]. The total fields in medium I can be described as the superposition of the incident and the diffracted field E_i and E_d:

$$E_I = E_i + E_d. \tag{1.29}$$

Due to the linearity of the Helmholtz equation both constituent fields E_i and E_d must obey the Helmholtz equation. Since there is no field in medium II, the boundary condition at the reflective grating surface requires the field to disappear. The goal is now to identify a periodic part E_p in the diffracted field. This can be achieved with the following approach:

$$E_d(x,z) = \underbrace{E_d(x,z)\, e^{-ik_{i;x}x}}_{E_p(x,z)} \underbrace{e^{ik_{i;x}x}}_{e^{i\phi(x)}}, \tag{1.30}$$

where the position-depended phase term $e^{i\phi(x)}$ can be recognized as the influence of the tilted incident light.

[2]Note that the letter i is never used as an index in this work! Depending on the context it can easily be identified as the imaginary number or as a subscript, which indicates the affiliation with the incident field.

In order to show, that E_p is truly periodic in x, the following equations must hold:

$$E_p\left(x+P_x,z\right) \stackrel{!}{=} E_p\left(x,z\right) \tag{1.31}$$

$$E_d\left(x+P_x,z\right)e^{-ik_{i;x}(x+P_x)} \stackrel{!}{=} E_d\left(x,z\right)e^{-ik_{i;x}x} \tag{1.32}$$

$$\Rightarrow E_d\left(x+P_x,z\right)e^{-ik_{i;x}P_x} \stackrel{!}{=} E_d\left(x,z\right) \tag{1.33}$$

However, the final statement is only true if the solution of the diffraction problem is unique. This is in fact a quite difficult task to prove, but can be reviewed in [19]. Now, one must only show, that the left side of Eq. 1.33 is also a solution of the diffraction problem and both sides of the equation must be equal. The diffraction problem is satisfied, if the left side is a solution of the Helmholtz equation and also obeys the boundary condition. Obviously, the Helmholtz equation is satisfied by its linear nature, since E_d is a solution. The boundary condition can be derived from the periodicity of the the incident plane wave:

$$E_i\left(x+P_x,z\right) = E_0 e^{i(k_{i;x}(x+P_x)+k_{i;z}z)} = E_i\left(x,z\right)e^{ik_{i;x}P_x}, \tag{1.34}$$

which allows the following substitutions at the boundary $(z = f\left(x\right))$,
where $E_i + E_d = 0$:

$$E_d\left(x+P_x,f\left(x\right)\right)e^{-ik_{i;x}P_x} = -E_i\left(x,f\left(x\right)\right)e^{ik_{i;x}P_x}e^{-ik_{i;x}P_x} \tag{1.35a}$$

$$= E_d\left(x,f\left(x\right)\right). \tag{1.35b}$$

Therewith, the initial statement, that the diffracted field E_d consists of a periodic function E_p and an additional position-depended phase term $\exp\left(ik_{i;x}\left(x\right)\right)$, which stems from the tilted incident field, is proven. The diffracted field is therefore called quasi-periodic and can also easily be generalized for two-dimensional gratings using the descriptive surface function $f\left(x,y\right) = f\left(x+P_x,y+P_y\right)$.

The lateral coordinates x and y can now be separated using the Fourier approach and expanding the periodic part E_p of the diffracted field into plane waves along its periodic directions:

$$E_d\left(x,y,z\right) = E_p\left(x,y,z\right)e^{i(k_{i;x}x+k_{i;y}y)} \tag{1.36a}$$

$$= \left[\sum_{m,n} S_{p;m,n}\left(z\right)e^{2\pi i\left(\frac{m}{P_x}x+\frac{n}{P_y}y\right)}\right]e^{i(k_{i;x}x+k_{i;y}y)} \tag{1.36b}$$

$$= \sum_{m,n} S_{p;m,n}\left(z\right)e^{i(k_{x;m}x+k_{y;n}y)} \tag{1.36c}$$

Here, $S_{p;m,n}$ is the still z-dependent Fourier coefficients of the electric field E_p and the

directions of the discrete diffraction orders are defined as

$$k_{x;m} = k_{i;x} + m\frac{2\pi}{P_x} \quad \text{and} \tag{1.37a}$$

$$k_{y;n} = k_{i;y} + n\frac{2\pi}{P_y} \tag{1.37b}$$

with mode indices m and n ranging from $-\infty$ to $+\infty$. Equations 1.37a and 1.37b can also be identified as *grating equations*[3]. (For example, canceling $n_I\frac{2\pi}{\lambda}$ transforms Eq. 1.37a to $s_x = s_{i;x} + \frac{m\lambda}{n_I P_x}$ with n_I as the refractive index of the incident medium. Also see Eq. 1.14.)

Substituting E_d into the Helmholtz equation $\nabla^2 E(\mathbf{r}) + k_0^2 E(\mathbf{r}) = 0$ gives

$$\sum_{m,n}\left[\left(k^2 - k_{x;m}^2 - k_{y;n}^2\right) S_{p;m,n}(z) + \frac{\partial^2}{\partial z^2} u_{s;m,n}(z)\right] e^{i(k_{x;m}x + k_{y;n}y)} = 0. \tag{1.38}$$

Due to orthogonality of the complex Fourier series[58] this must be true for every value of m and n. Using Pythagoras' theorem (cf. Eq. 1.16) this results in the second-order differential equation

$$\frac{\partial^2}{\partial z^2} S_{p;m,n}(z) + k_{z;m,n}^2 S_{p;m,n}(z) = 0. \tag{1.39}$$

Equation 1.39 meets the characteristics of an eigenvalue problem: $g(S_p) = \lambda \cdot S_p$, where λ is the eigenvalue and S_p is the eigenfunction of the linear operator g. This is a second-order differential equation with constant coefficients, which can be solved through its characteristic equation:

$$\nu^2 - \lambda = 0 \tag{1.40}$$

The characteristic equation 1.40 has two simple zeros at $\nu_1 = +\sqrt{\lambda}$ and $\nu_2 = -\sqrt{\lambda}$, and due to the derivative nature of the exponential function, $\{\exp(\nu_l z) \mid l = 1, 2\}$ forms a basis of the solution space for the differential equation. Thus, the general solution to Eq. 1.39 can be written as a linear combination of these independent solutions:

$$S_{p;m,n}(z) = c_{1;m,n} e^{+\sqrt{\lambda_{m,n}}z} + c_{2;m,n} e^{-\sqrt{\lambda_{m,n}}z}. \tag{1.41}$$

c_1 and c_2 are complex coefficients that are only determined by the exact boundary conditions and Sommerfeld's radiation condition even requires $c_1 = 0$, since a positive real number of

[3] The grating equation $s_x = s_{i;x} + \frac{m\lambda}{n_I P_x}$ also shows that for larger structures (larger values of P_x) the higher diffraction order move closer together until they are indistinguishable from the 0th order for macroscopic structures and can be described by geometrical optics.

λ would result in an unphysical growth of the first term as $z \to \infty$. The remaining second term shows the exponential characteristic of a plane waves. And by substituting it into the differential equation 1.39 one can in fact identify the eigenvalues with the z-component of the wavevector:

$$\lambda_{m,n} \equiv -k_{z;m,n}^2 \Rightarrow \sqrt{\lambda_{m,n}} \equiv \pm i k_{z;m,n} \tag{1.42}$$

Since the initial grating was assumed to be a perfect reflector, the diffracted field must be equal to the reflected field. Thus, the Fourier coefficients $S_{p;m,n}$ can be identified as reflective coefficients $R_{m,n}$ and the field in medium I can be expressed as

$$E_I(x,y,z) = E_0 e^{i\mathbf{k_i r}} + \sum_{m,n} R_{m,n} e^{i(k_{x;m}x + k_{y;m}y + k_{z;m,n}z)}. \tag{1.43}$$

This is the so called Rayleigh expansion, where the field is composed of a set of discrete plane waves, also known as modes with different directions of propagation. For tilted incidence, the result is a pseudo-periodic field and the modes of the diffracted field are called Floquet harmonics.
For real, non-perfectly conducting gratings, which allows light transmission, the Rayleigh expansion also holds for the field in the transmission region behind the grating[1]. Inside the grating region, however, the eigenvalue problem can not be resolved as easily (cf. Eq. 1.42), since the complex permittivity is a function of x and y. In fact, this limitation goes even further and restricts the validity of the Rayleigh expansion to smooth gratings with small modulation depth[4]. The reason lies in the fact that the diffracted outside fields are always described by plane waves of only one *direction* (meaning either forward or backward)[131]. And it is plausible, that this is not sufficient to describe the scattering at deep grooves and exotic grating profiles.

The rigorous coupled-wave analysis basically uses the same approach as described here. However, it circumvents its limitations by partitioning the grating region into a stack of inhomogeneous layers. And while the fields in the homogeneous outside regions are actually described by Rayleigh expansion, the eigenvalue problem in the inhomogeneous layers of the grating region is solved numerically. In addition the general diffraction problem cannot be separated into the two distinct forms of polarization (cf. Sec. 1.2.1) and therefore, the numerical approach also needs to account for the coupling between the electric and magnetic components in a larger system of differential equations.

[4] Petit and Cadilhac[52] and Millar[86] were able to prove that the Rayleigh expansion is invalid for $h \cdot 2\pi/P_x > 0.448$, where h is the groove depth of the grating

1.2.2 Scalar methods of diffraction theory

The angular spectrum method (ASM), which was described in Sec. 1.1.6, is an important method to solve the diffraction problem in homogeneous media. The required knowledge about the field amplitude at $z = 0$ is also the characteristic of the so called Dirichlet boundary condition. If, on the other hand, only the first derivative of the field $\frac{\partial}{\partial z}\mathbf{E}(\mathbf{r}_\perp, 0)$ is known, Eq. 1.20 transforms into the angular spectrum method under Neumann boundary conditions. Both forms of the ASM equally satisfy not only the vector form of the wave equations, but also the scalar Helmholtz equation (cf. Sec. 1.15). And in fact they are used as such in classical diffraction theory, which considers any electromagnetic field to be scalar. Moreover, the representation of the fields is not limited to plane waves and spherical waves according to the Huygens–Fresnel principle were even considered first. If the decomposition of the ASM is performed with Weyl's definition of spherical waves $\left(E(\mathbf{r}) = e^{i\mathbf{kr}}/|\mathbf{r}|\right)$, the two described boundary conditions lead to Sommerfeld's diffraction integrals of the first and second kind respectively - and to the diffraction integral of Kirchhoff, which is the arithmetic average of the two. Other important approaches of classical diffraction theory are the far field approximation by Frauenhofer, the paraxial approximation by Fresnel (cf. Eq. 3.31) and the Debye integral for the representation of focused fields. A detailed description of these methods can be found in almost any standard text book, which covers diffraction theory (e.g. Born and Wolf[11, Chap. 8, 13]). Common scalar methods that describe structure light interaction are the thin element approximation (TEA)[71], the beam propagation method (BPM)[35] and wave propagation method (WPM)[17].
According to Pommet et. al[104] scalar diffraction theory is valid (with deviations below 5%) for the analysis of single-level diffractive phase elements down to a minimum feature size of 14 wavelengths, while the accuracy declines as the index of refraction is lowered and as the fill factor deviates from 0.5. A distinct limit, however, cannot be drawn, since it depends on the geometry and composition of the diffractive elements.
Since the rigorous coupled-wave analysis, as a numerical approach to solve the full, vector-based diffraction problem, is elaborated (cf. Chap. 2), extended (cf. Chap. 3) and implemented in this work, further discussions on classical diffraction theory are spared here, even though the topic is by all means relevant both historically and practically. The reader is instead referred to the literature cited above.

1.2.3 Rigorous methods of diffraction theory

On the scale, where the diffractive structure size is of the same order of magnitude as the wavelength, the vector nature of the electromagnetic fields plays an important role and an appropriate model needs to obey the close interdependence between the electric and magnetic field components, which follow from Maxwell's equations. This is the regime of the rigorous methods. And they are characterized by the obedience of exactly these two aspects. Despite the difficulty of a general solution of the rigorous diffraction problem this regime is particularly interesting, since optimized geometries in the near field can lead to valuable properties like high dispersion or light concentration in a single diffraction order. It is also very significant for this work due to strong resonances, which can be induced. Utilized in

the right way these can lead to high local absorption and thereby improve the performance of photo detecting devices (cf. Chap. 6).

The term rigorous only refers to the full vectorial description of the field and Maxwell's equations. Despite the promising name, there are in fact several constraining assumptions, that can be considered as approximations like an indefinite periodicity of the diffractive structures, steady state behavior, linearity and isotropy of the optical material, smooth surfaces[5] and the absence of all effects that can only be explained by quantum eletrodynamics (QED). Furthermore, most practical diffraction problems can only be solved by numerical methods, which often struggle with bad convergence and instabilities and may thereby further reduce the accuracy of the results.

As mentioned before, the vector character of the wave equation becomes particularly important in the presence of inhomogeneities of the material properties, since it leads to an interdependence of the vector field components (cf. Sec. 1.1.4). This prevents a direct solution of the wave equations and its decoupling is therefore the main challenge of the rigorous diffraction problem. Analytical solutions to Maxwell's equations are only known for few cases. For one there are perfect conductors like Sommerfeld's diffraction at the perfectly conducting half-plane[117], but most solutions are obtained by the method of separation of variables. In many cases, the inhomogeneities in permittivity and permeability are confined to boundaries between homogeneous materials. However, these jump discontinuity in the material properties are problematic, because the partial derivatives in Maxwell's equations are not defined here. The method of separation of variables is a way to circumvent this issue. The idea is to transform Maxwell's equations into a coordinate system, which follows the geometry of the diffractive object, expand the electromagnetic fields into orthogonal fundamental modes and solve the expansion coefficients at the object's boundary by separately applying Maxwell's continuity conditions[60, p. 31].
Unfortunately, the wave equation is only known to separate in eleven coordinate systems including euclidean 3-space and spherical. Here, two important theories should be mentioned: The first one is the multilayer matrix theory[14, 21], which in fact does not solve a diffraction problem, but still provides an analytical and rigorous solution to Maxwell's equations for the interaction with a stack of different homogeneous layers. Applications include selective wavelength filters and anti-reflective coatings. The second one is the Mie theory[85], which addresses spherical diffractive objects and is used for the study of nanoparticles.

Since the number of numerical methods is vast, this discussion does not try to provide a complete list, but only addresses the most common ones. It also does not try to compare the methods in terms of overall performance. Since each of the methods was initially designed for a specific problem all methods have different strength and drawbacks depending on the application. However, there are some comparing studies, where domains are overlapping. A performance comparison between various methods for a basic, non-periodic scatterer was done by Lalanne et al.[69]. Some other references can be found in the text. Very detailed overviews over different numerical methods for the rigorous diffraction problem can also be found in Kleemann's work[64] and also in the books of Nevière and Popov[97, Chap. 1.1.3] and Loewen and Popov[Chap. 10.5-10.10][79].

[5] Imperfections due to manufacturing are assumed to be invisible to the light.

Due to the difficulties that arise in the description of the fields at the material boundaries, the following overview will distinguish between those methods whose elementary solutions cross material interfaces and those who manage to avoid it (cf. Fig. 1.4). They will be referred to as *intersecting* and *non-intersecting* methods respectively.

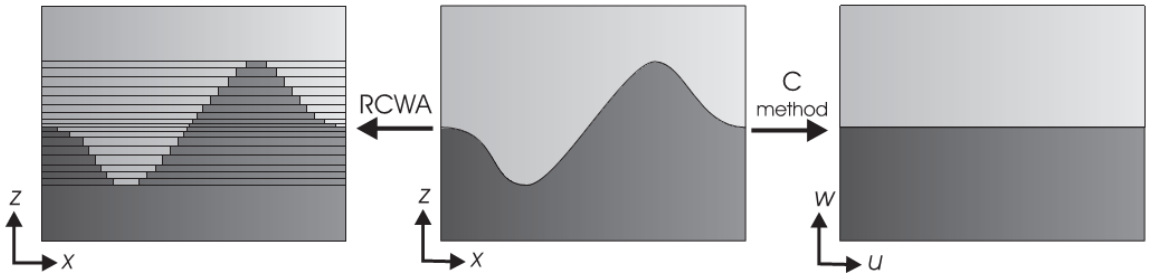


Figure 1.4: Example of an intersecting and a non-intersecting method: The schematic representation shows the way the RCWA and C method remove the z dependency. In case of the RCWA the lateral dependencies remain and both fields and material properties are expanded into Fourier base function *across* material boundaries.
Source: N.P. van der Aa. "Diffraction grating theory with RCWA or the C method", *Progress in Industrial Mathematics at ECMI 2004*. Springer, 2006, pp. 102

Intersecting methods

The differential method and the rigorous coupled-wave analysis both separate the z-dependence[6] of the material properties by a segmentation of space into plane layers along z (cf. Fig. 1.4). Thus, in each layer the permittivity (and also the permeability if considered) is constant in z, but may vary periodically along the lateral directions of the grating. The material properties as well as the electromagnetic fields inside the grating are Fourier expanded in lateral direction and across potential jump discontinuities. At these boundaries the approximated material properties and the fields exhibit complementary jumps, which lead to numerical artifacts and instabilities of the methods (cf. Sec. 2.5). The accepted solution is the appropriate usage of the so called *inverse rule* depending on the orientation of the boundary and the polarization in compliance with Maxwell's boundary conditions (cf. Sec. 1.1.7). The Fourier expansion of the Maxwell equations in lateral direction leads to a coupled system of second order partial differential equations, which is numerically integrated by the differential method. The method was first presented in 1974 by Nevière[98]. A detailed description of the method is given by Maystre[82, Chap. 4], [105]. Interestingly, the differental method is not confined to Fourier base function and can thereby be adjusted to the diffraction problem. Examples are cylindrical Bessel functions[10] for diffraction by a single aperture or spherical functions[119] for arbitrarily shaped single objects.

The rigorous coupled-wave analysis (RCWA) starts with the same steps as the differential method, but solves the coupled system of second order partial differential equations through an eigenvalue approach and thereby transforms the coupled Fourier modes into a

[6] z is the preferred propagation direction of the incident light, normal to the grating surface (cf. Fig. 1.3)

set of uncoupled eigenmodes. In the RCWA and in the differential method, the fields in the outside region are described by a Rayleigh expansion and the expansion coefficients can be solved by applying Maxwell's continuity conditions at the boundaries. Due to a continuous development over the years the RCWA is able to provides stable solutions for a wide range of diffraction problems and thus has become one of the most common methods to solve grating diffraction problems. Noteworthy, are the use of perturbation theory to speed up the solution of the eigenvalue problem by Edee et al.[32] and the replacement of the Fourier basis functions by Legendre polynomials(Khavasi et al.[62]) or by Chebyshev polynomials(Yeh et al. [135]) in order to improve the approximation of steep gratings grooves and thereby avoid the inherent numerical instabilities of the RCWA. A detailed discussion and further extension of the RCWA method is given in Chap. 2 and 3, respectively.

The finite element method (FEM) was initially developed around 1960 to solve partial differential equations for structural engineering, but since has been applied in a wide range of fields[57]. The first implementations for diffraction gratings was published by Delort et al.[27]. The FEM samples Maxwell's equations in real coordinate space on a mesh of triangles or tetrahedrons of dynamic size. Associated with the mesh are locally defined trial functions on each sampling point, which approximate the field. The continuity conditions according to Maxwell's equations are applied wherever material boundaries coincide with a discretization point. The weights of every trial function is then determined by a linear system of equations. The region in between the discretization points is interpolated. Since the method is applied in real space, problems with high spatial frequencies, which might occur at steep grating sloped, do not occur. The FEM is also able to model non-periodic structures and is a straightforward approach even for sophisticated grating geometries. However, highly conducting gratings require very high density discretization meshes, which leads to long computation times.

The finite difference time domain method (FDTD) was first published by Yee[134] and extended to infinite gratings by Ichikawa[53]. The method discretized both real space and time and replaces the partial derivatives of Maxwell's equations by finite differences. Since also time is discretized, the method is especially interesting for non-harmonic processes, like light pulses. Due to the four-dimensional discretization and since Maxwell's equations are solved for both electric and magnetic fields, the method can require a significant amount of computation time memory storage and might therefore be less adequate for typical harmonic grating problem.

Non-intersecting methods

The Rayleigh expansion was discussed in Sec. 1.2.1 as simple way of describing pseudo-periodic fields by set of plane waves. Although the approach is only strictly valid in the outside region of a grating, it has a significant value as an integral part of other theories like the RCWA. The method is also described in great detail by Petit[103], Neviere[97, Chap. 1.2.3-5] and Hench et al.[48]. A discussion about the validity for gratings (also known as Rayleigh hypothesis) can be found in the works of Petit and Cadilhac[52], Millar[86], Wauer et al.[131] and in the book of Loewen and Popov[Chap. 10.5-10.10][79].

The C-method of Chandezon[20] is based on the idea that the Rayleigh expansion would be valid everywhere, except at the grating interface, if the interface were flat[1]. Hence, it uses a coordinate transformation to straighten the grating modulation to a flat surface. This way the differential Maxwell's equations can be integrated separately in each medium. However, the algorithms leads to an eigenvalue problem due to varying coefficients in the differential equation, which needs to be solved numerically. Furthermore, Maxwell's equations cannot be separated into independent cases for TE and TM polarization. An advantage compared to the RCWA is that the C-method does not rely on a staircase approximation for continuous grating profiles. A comparison between the C-method and the RCWA is given by Vallius[127] and Van der Aa[1].

The classical modal method separates space into rectangular regions of homogeneous material properties and consider each of these regions as lateral waveguides[26]. The containing fields are described by a set of harmonic waveguide modes, whose expansion coefficients are solved by applying Maxwell's continuity conditions at the boundaries. The classical modal method can also be seen as a special case of the RCWA, which also allows a continuous modulation of the material properties in lateral direction.

The integral method was one of the first approaches to provide a solution to the rigorous diffraction problem. The method was first developed for perfectly conducting gratings by Petit[102] in 1965. Later works by Kleemann et al.[65][64] allowed the study of more general surface profiles with vertical walls and undercuts by using the boundary integral equation system method with parametrization(IESMP). The method also avoids the crossing of the material boundaries. Instead it considers a surface current on both sides of the boundary, which is induced by the incident wave. Every point along the boundary above and below the surface is thereby considered a secondary source and the total current is obtained by an integral, which can be solved by distribution theory. Green's theorem is then used to derive the field inside the boundaries from the field and its normal derivative on the boundary curve. The method is able to handle highly conducting gratings even with deep gratings and arbitrary profile, but the advantage comes at the cost of complex mathematics and also high computational resources both in time and memory storage[79, Chap. 10.7]. A detailed description of the integral method is given by Maystre and Popov[82, Chap. 3], [83].

The multiple multipole method (MMM) was first suggested by Singer[115]. The method divides space into contiguous domains of homogeneous material properties. In each domain the field is expanded into a linear combination of analytical solutions of the Maxwell equations. Since the introduction of the generalized multipole technique[44], the method is flexible in the choice of these basis function and can be adjusted to the problem. However, the most common ones are still multipolar functions, which can be found by *applying the appropriate differential operators to a solution of the scalar Helmholtz equation, separated in spherical coordinates*[94]. These solutions represent point sources, which are placed inside the diffracting structure and close to the domain interface in order to radiate in all directions and especially beyond the interface. The method is therefore also referred to as ficticious sources method[124]. The total field is then generated by the exciting source and all ficticious sources. The weights of the analytical basis functions are determined by minimizing the errors with regard to the boundary condition. This also leads to a special feature of the approach, which is the ability to estimate the quality of the method. Since Maxwell's continuity conditions are only satisfied on a discretized boundaries, whereas the basic functions are analytical solution, the method is also called a semi-analytical approach[43]. The biggest problem of the method is the difficulty of a proper placement of the ficticious sources, which is not intuitive and requires high expertise of the user. However, attempts for automatic placements[94] do exist. It is evident that the method is not limited to periodic gratings.

The discrete dipole approximation (DDA) also known as coupled dipole method was first developed by DeVoe[28, 29] in 1964 to study the optical properties of molecular aggregates. The method approximates the substance of a diffractive object with polarizable points. Each such point acquires its dipole moment in response to the incident field and interacts with every other one through its electric field. For a finite number of oscillating point dipole moments in monochromatic excitation the self-consistent solution to the scattering problem can be found analytically. Thus the only approximation in the DDA is the discretization of the object into a finite number point dipoles. Absorption and scattering throughout the simulation domain and across the material bounds can finally be derived from the polarization of the point dipoles[30]. The method is applicable[125] but clearly not limited to periodic gratings and can be extended to anisotropic materials and materials with nonzero magnetic susceptibility. The method lately draw some attention due to a fast implementation of its field calculations[38]. The code is public domain[31].

The inevitable but significant increase in computation time and memory usage of the rigorous methods has also led to the development of semi-rigorous methods like the vectorial beam propagation method (VWPM)[36] and the vectorial thin element approximation (VTEA)[61], which try to close the gap between classical and rigorous methods by extending the scalar methods to vector fields.

Chapter 2

The standard RCWA algorithm

The rigorous coupled-wave analysis (RCWA) is a well-established method for the simulation of diffraction and near field effects and is also the basis of this work. As such it serves as the main simulation method and is further developed in chapter 3. After a historical review, a detailed description and discussion of the RCWA algorithm is the purpose of this chapter.

2.1 Historical review

As already mentioned, even rigorous methods are not without physical approximations and even less without unlimited validity. In fact the rigorous coupled-wave theory was first developed under many restrictive assumptions, some of which have been overcome over the years, while others still remain. Without claiming completeness, some important developments are listed in the following:

In 1914 Darwin[25], who studied the diffraction of X-rays in crystals, first came up with the idea to treat Bragg diffraction with the concept of coupled-waves. This was adopted by Burckhardt[18] in 1966, who analyzed diffraction of single plane waves in holographic gratings inside thick photographic emulsions. It was observed in experiments, that in those setups diffraction behaves similar to Bragg diffraction as diffracted beams reach their maximum intensity at the Bragg angle. Hence, thick films could not be treated with the Fresnel-Kirchhoff integral and thin element approximation. Instead, Burckardt used a coupled-wave method based on Floquet theorem to study gratings with a sinusoidal modulations of the refractive index. In 1969 also Kogelnik[66] developed a coupled-wave theory for sinusoidal gratings. However his theory was restricted to the two first diffraction orders only. In 1973 Kasper[59] extended Burckardt's theory to general periodic one-dimensional gratings. He also introduced the complex dielectric constant[1] (cf. Sec. 1.1.4) for the treatment of absorbing gratings.

Since rigorous treatments of the diffraction problem showed steadily growing importance, in 1981 Moharam and Gaylord[88] published a condensed and more transparent formulation with the ability to numerically analyze a wide variety of one-dimensional grating problems. Their approach also allowed the calculation of oblique incidence in a plane normal to the grating grooves. It was also this paper, which established the name rigorous coupled-wave analysis (RCWA) for the method. Shortly after, Moharam and Gaylord extended the the-

[1] Today the dielectric constant is commonly referred to as relative permittivity (cf. Sec. 1.1.2).

ory to surface-relief gratings[87] by introducing a multi-layer staircase approximation (cf. Sec. 2.8) and also to conical mounting[89] (cf. Fig. 1.3b). Two-dimensional grating structures were first studied by Bräuer and Bryngdahl in 1993[12]. 1995, Moharam and Gaylord presented a revised formulation of the RCWA[91] (cf. Sec. 2.2). In it, former numerical instabilities, which caused the appearance of unphysical, exponentially growing fields (also called anti-evanescent waves), had been eliminated by appropriate normalization and substitution. In the same year, 1995, Moharam and Gaylord presented yet another significant work[92] introducing a numerically stable treatment of surface-reliefs and multi-level gratings, which again faced the problem of exponentially growing fields with an elegant normalization scheme (cf. Sec. 2.8.3). This method, which they called the enhanced transmission matrix approach (ETMA) was later shown[123] to be more efficient than previous approaches like the S-matrix, the R-matrix approach[72, 73] or the straight forward T-matrix approach (cf. Sec. 2.8.1 - 2.8.3).

The subsequent developments particularly concerned a convergence problem, which is still topic of discussion today. Due to the Fourier expansion of the electromagnetic fields and the permittivity function, the representation of binary grating steps provoke the occurrence of a Gibbs phenomenon. In one-dimensional grating configurations this concerns the x component of the electric field, if the incident light is TM polarized, and causes a significant drop in convergence performance. In 1996, independently of each other, Lalanne and Morris[68] as well as Granet and Guizal[41] empirically discovered a way to overcome this problem for the one-dimensional case. Today, the method is known as the Li's factorization rules (cf. Sec. 2.5.2), whose author provided a detailed description[77] on their use and also added a formalism for the treatment of two-dimensional crossed gratings[76]. In 2009 Schuster[113] introduced the concept of normal vector fields to the RCWA (cf. Sec. 2.5.5), based on the idea of Popov and Nevière[107]. This work can also be seen as a generalization of Li's rules for two-dimensional gratings: While previously slanted grating contours had to be approximated by a zigzag contour in order to apply Li's rule, this restriction was overcome here. However, the new approach requires the continuation of the normal vector fields in the complete simulation area, which is not trivial for complex geometries. Another important extension of the RCWA is the concept of perfectly matched layers (PML) by Silberstein and Lalanne in 2001[114], which even enables the simulation of non-periodic structures. Again inspired by the work of Popov and Nevière[107], 2003 Li included anisotropic media and locally dependent permeability to the scope of the RCWA[74].
While all above studies only consider diffraction for plane wave incidence, the present work aims to extend the RCWA to a natural treatment of localized input fields (cf. LIFRCWA in Chap. 3) like Gaussian beams. This, for example, allows the rigorous analysis of waveguide coupling. Besides, near fields (cf. Chap. 4) and local absorption (cf. Chap. 5), which derives from the rigorous calculation, is investigated for the use of component optimization.

2.2 Definition

In this work, the term standard RCWA refers to the formulation of Moharam and Gaylord in 1995[91] for gratings with periodicity in two dimensions. Therein, all media are assumed to be linear, isotropic, non-conducting and having a permeability equal to that of the vacuum. The incident wave is assumed to be a monochromatic plane wave with harmonic time dependence. In contrast to previous works, however, the following formulation is slightly improved in some points to enhance clarity: For one, the diffraction coefficients are obtained using only a single global Cartesian coordinate system instead of a local coordinate system for every diffraction order. Besides, the eigenvalue problem is applied to the electric field components rather than the magnetic ones and the notation is slightly adjusted to be naturally compatible with the LIF extension of chapter 3. As such the standard RCWA will later serve as a reference for further extensions and discussions.

2.3 Geometry

The following section describes the geometry of a typical three-dimensional diffraction problem for the RCWA.

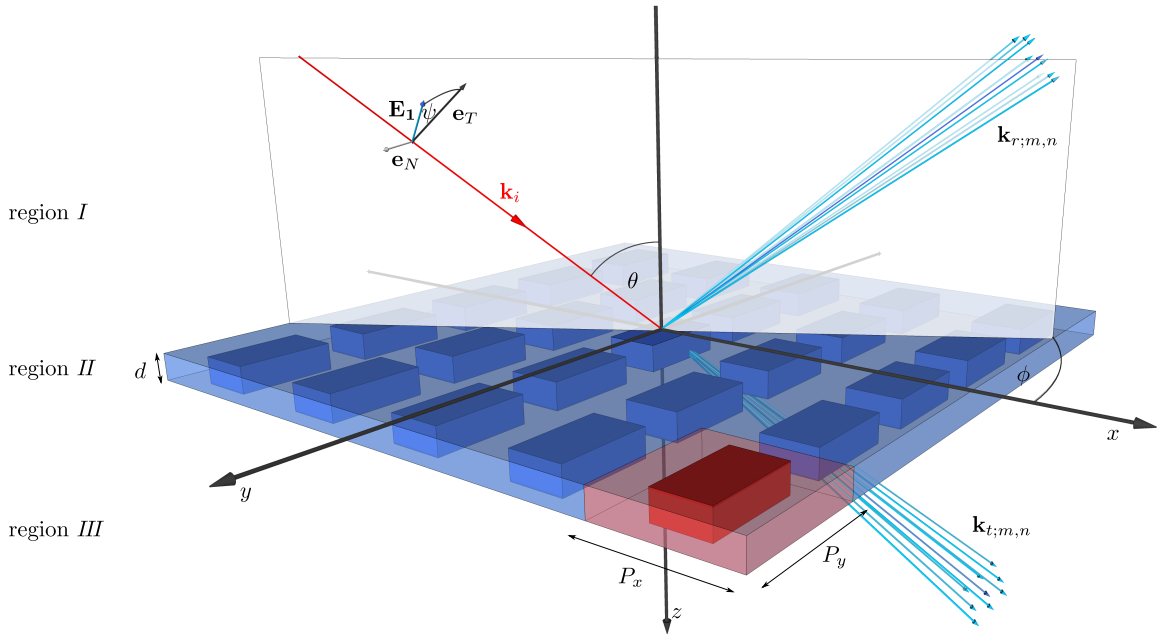


Figure 2.1: Geometry of the three-dimensional diffraction problem at oblique incidence

2.3.1 Simulation area

The simulation area is defined in Cartesian space and divided into three regions along z with planar interfaces in between. Region *I* and region *III* are also known as superstrate and substrate. They are assumed to be homogeneous with constant refractive index n_I and n_{III} respectively. Consequently, diffraction only takes place in the inhomogeneous region *II* in between.

2.3.2 Diffractive structure

A diffractive layer in region *II* is characterized by its thickness d and a function $n_{II}(x, y)$ of its refractive index. The latter may vary with constant periodicity P_x and P_y in lateral directions, but remains constant in the direction of z. The result can be a two-dimensional binary grating, as illustrated in Fig 2.1, but also gradient index structures are allowed. In the standard RCWA region *II* contains only one such diffractive layer. Extensions to multilayer gratings are discussed in Sec. 2.8.

Since in most optical applications the permeability of the structure can be neglected (refractive index $n = \sqrt{\epsilon\mu}$ with relative permeability $\mu = 1$), the isotropic refractive index of the grating can be fully described by a local permittivity function $\epsilon_{II}(x, y)$. Due to its periodic definition $\epsilon_{II}(x, y)$ can be expressed by a Fourier expansion as it is used in the following RCWA algorithm:

$$\epsilon\,(x, y) = \sum_{m,n} \epsilon_{m,n} e^{2\pi i\left(\frac{m}{P_x}x + \frac{n}{P_y}y\right)} \tag{2.1}$$

To improve readability here and in the following, ϵ_{II} will be referenced without further mention of the index *II*.

When using Fourier expansion in numerical simulations, a truncation of the infinite series to a finite sum of terms is always inevitable. The limits of the resulting symmetrically truncated partial sums are defined by the mode count $M = 2M_0 + 1$ and $N = 2N_0 + 1$ for each dimension respectively. Thus, the individual mode indices m and n range in symmetrical intervals $\{-M_0\,.\,+M_0\}$ and $\{-N_0\,...\,+N_0\}$. The Fourier coefficients $\epsilon_{m,n}$ are determined by the inverse Fourier transform and can either be obtained analytically or numerically. In case of an analytic approach, the solution to the Fourier integral

$$\epsilon_{m,n} = \frac{1}{P_x P_y} \int_{P_x} \int_{P_y} \epsilon\,(x, y)\, e^{-2\pi i\left(\frac{m}{P_x}x + \frac{n}{P_y}y\right)} dx\, dy \tag{2.2}$$

needs to be known and thus the computational effort almost dissolves. Moreover, the coefficients can be parameterized and then controlled directly in the Fourier domain. For any other arbitrary periodic and discrete (or discretized) space function, for which the Fourier integral is either unknown or not available, the coefficients can still be calculated using discrete Fourier transform:

$$\epsilon_{m,n} = \frac{1}{N_x N_y} \sum_{p,q} \epsilon\,(x_p, y_q)\, e^{-2\pi i\left(\frac{m}{P_x}x_p + \frac{n}{P_y}y_q\right)}. \tag{2.3}$$

2.3.3 Incident field

For now, the setup is illuminated by a single plane wave in region I. The angle of incidence θ is defined with respect to the surface normal and the azimuth angle of the plane of incidence is ϕ (cf. Fig. 2.1). A local coordinate system spanned by the vectors $\mathbf{e}_T$ and $\mathbf{e}_N$ (tangential and normal to the plane of incidence) is used to describe the field's polarization state[2]. Here, ψ is defined as the angle between the polarization vector of the electric field $\mathbf{E_1}$ and the polarization basis vector $\mathbf{e}_T$ in the plane of incidence. $\mathbf{k}_i = n_I k_0 \mathbf{s}_i$ is the wave vector of the electromagnetic incident field and points in the propagation direction of the wave.

Recalling the definition of the vectorial plane wave (cf. Eq. 1.12), the incident field can be described by[3]

$$\mathbf{E}_i(\mathbf{r}) = E_0 \cdot \mathbf{E_1} \cdot e^{i\mathbf{k}_i \mathbf{r}}. \tag{2.4}$$

E_0 is a scalar amplitude an usually set to one in simulations. Substituting $\mathbf{E_1}$ according to the definition given in 1.26 and the normalized wave vector $\mathbf{s}$ according to Eq. A.1 leads to a full description of the incidence field with the three defining angles θ, ϕ and ψ for propagation and polarization:

$$\mathbf{E}_i(\mathbf{r}) = \frac{1}{s_{i;\perp}} \begin{pmatrix} cos(\psi)s_{i;x}s_{i;z} - sin(\psi)s_{i;y} \\ cos(\psi)s_{i;y}s_{i;z} + sin(\psi)s_{i;x} \\ -cos(\psi)s_{i;\perp}^2 \end{pmatrix} e^{ik_0 n_I (s_{i;x}x + s_{i;y}y + s_{i;z}z)} \text{ for } s_{i;\perp} \neq 0 \tag{2.5a}$$

$$= \begin{pmatrix} cos(\psi)cos(\theta)cos(\phi) - sin(\psi)sin(\phi) \\ cos(\psi)cos(\theta)sin(\phi) + sin(\psi)cos(\phi) \\ -cos(\psi)sin(\theta) \end{pmatrix} e^{ik_0 n_I (sin(\theta)cos(\phi)x + sin(\theta)sin(\phi)y + cos(\theta)z)} \tag{2.5b}$$

2.4 Diffracted fields

Since the diffracted fields of a periodic grating structure are also assumed to follow the same periodicity (cf. Sec. 1.2.1), a key idea of the RCWA algorithm is the use of a Fourier expansion approach for the entire electromagnetic field inside and outside the grating region in order to dissolve the spatial derivatives in Maxwell's equations (cf. Eqs. 1.1a and 1.1b). In general, however, the electromagnetic fields do not show strict periodic form along the

[2] The standard RCWA only provides for a single incident plane wave and thus only allows linear (and circular) polarization of the incident light. An extension of the RCWA algorithm to localized input fields, which provides more flexibility in the polarization is introduced in Chap. 3.

[3] Note again that the letter i is never used as an index in this work. Depending on the context it can easily be identified as the imaginary number or as a subscript, which indicates the affiliation with the incident field.

grating surface. In fact, according to the Floquet-Bloch theorem, the general solution of the grating diffraction problem involves pseudo-periodic fields as shown in Sec. 1.2.1. This particularly applies to tilted incidence and is described by an additional phase term, when expanding the fields into so-called pseudo-Fourier series. In the limit of normal incidence, the phase term disappears and the field becomes strictly periodic.
The (Pseudo-)Fourier expansion of the electromagnetic field provides an equivalent representation by superposition of plane waves and marks the transition from a continuous (pseudo-)periodic spatial field to a discrete representation in Fourier space, in which the RCWA algorithm operates.

2.4.1 Diffracted field outside the grating

Outside the grating, the Pseudo-Fourier expansion of the electric field is carried out along the lateral coordinates (x and y), whereas the normal dependence (z) is realized as a simple wave propagation term. The result is a Rayleigh expansion of the field (cf. Sec. 1.2.1).
In region I the electric field $\mathbf{E}_I(\mathbf{r})$ is formed by the superposition of the incident wave (forwards propagating) and the reflected waves (backwards propagating), while in region III the electric field $\mathbf{E}_{III}(\mathbf{r})$ only consists of transmitted waves (forwards propagating):

$$E_{I;j}(\mathbf{r}) = E_{i;j}(\mathbf{r}) + \sum_{m,n} R_{j;m,n} e^{i\mathbf{k}_{r;m,n}\mathbf{r}} \tag{2.6a}$$

$$= E_0 E_{1;j} e^{i(k_{i;x}x + k_{i;y}y + k_{i;z}z)} + \sum_{m,n} R_{j;m,n} e^{i\left(k_{x;m}x + k_{y;n}y + \gamma_{I;m,n}z\right)} \tag{2.6b}$$

$$E_{III;j}(\mathbf{r}) = \sum_{m,n} T_{j;m,n} e^{i\mathbf{k}_{t;m,n}\mathbf{r}} \tag{2.6c}$$

$$= \sum_{m,n} T_{j;m,n} e^{i\left(k_{x;m}x + k_{y;n}y + \gamma_{III;m,n}z\right)} \tag{2.6d}$$

with $j = x, y, z$ and

$$k_{x;m} = k_{i;x} + 2\pi\frac{m}{P_x}, \qquad k_{i;x} = k_0 n_I s_{i;x} = k_0 n_I sin(\theta)cos(\phi), \tag{2.7a}$$

$$k_{y;n} = k_{i;y} + 2\pi\frac{n}{P_y}, \qquad k_{i;y} = k_0 n_I s_{i;y} = k_0 n_I sin(\theta)sin(\phi). \tag{2.7b}$$

In Eq. 2.6 $R_{j;m,n}$ and $T_{j;m,n}$ can be identified as the reflection and transmission coefficients of the (m,n)-th grating's diffraction order. $\mathbf{k}_{\rho;m,n}$ with $\rho = r, t$ defines the corresponding wave vector (cf. Fig. 2.1). Their lateral components in Eq. 2.7 consist of a harmonic part and an additional lateral shift $k_{i;j_\perp}$ in k-space with $j_\perp = x, y$. According to Snell's law they are equal throughout all regions. The normal components $k_{\rho;z;m,n}$, on the other hand, are

generally varying depending on the local refractive index according to the dispersion relation ($k = nk_0 = n2\pi/\lambda$, cf. Eq. 1.13b). Due to their special position, they are referred to with the symbol γ. For the outer regions the refractive index is assumed to be known and they can are defined as:

$$\gamma_{I;m,n} = -\sqrt{(k_0 n_I)^2 - k_{x;m}^2 - k_{y;n}^2} \quad \text{and} \tag{2.8a}$$

$$\gamma_{III;m,n} = \sqrt{(k_0 n_{III})^2 - k_{x;m}^2 - k_{y;n}^2}. \tag{2.8b}$$

2.4.2 Diffracted field inside the grating

Inside the grating the electromagnetic field is again expressed by a Pseudo-Fourier expansion along the lateral coordinates. However, since the grating is constant, but not periodic, the z-dependency remains unknown and thus the fields are only partially transformed into Fourier space:

$$E_{II;j}(\mathbf{r}) = \sum_{m,n} S_{j;m,n}(z) e^{i(k_{x;m}x + k_{y;n}y)} \tag{2.9a}$$

$$H_{II;j}(\mathbf{r}) = \frac{1}{Z_0} \sum_{m,n} U_{j;m,n}(z) e^{i(k_{x;m}x + k_{y;n}y)} \tag{2.9b}$$

$E_{II;j}$ and $H_{II;j}$ describe the electric and the magnetic spacial vector fields inside the grating, respectively. $S_{j;m,n}$ and $U_{j;m,n}$ are the corresponding Fourier coefficients with a remaining z-dependency. $Z_0 = \sqrt{\mu_0/\epsilon_0}$ is the impedance of the vacuum and its inverse is chosen as a prefactor of the magnetic field to cancel out different scaling factors of E- and H-fields in the following calculations.

Substituting Maxwell's equations

Assuming harmonic fields (cf. Eq. 1.8a and 1.8b) and a complex permittivity (cf. Eq. 1.9a), the first two Maxwell equations simplify to

$$\nabla \times \mathbf{H} = -i\omega\epsilon_0\epsilon\mathbf{E} \tag{2.10a}$$

$$\nabla \times \mathbf{E} = i\omega\mu_0\mu\mathbf{H}. \tag{2.10b}$$

For the later solution of the system of equations, the tangential components are most relevant due to their continuous nature at material boundaries (cf. Sec. 1.1.7). The z-components

can be eliminated without the loss of any information. Substituting the Fourier approaches of the permittivity and the fields leads to a differential vector equation, in which the lateral components of the electric and the magnetic field appear coupled:

$$\frac{\partial}{\partial z}\begin{pmatrix}\mathbf{S}_x\\ \mathbf{S}_y\\ \mathbf{U}_x\\ \mathbf{U}_y\end{pmatrix} = ik_0 \underbrace{\begin{pmatrix} 0 & 0 & \mathbf{K}_x[\![\epsilon]\!]^{-1}\mathbf{K}_y & -\mathbf{K}_x[\![\epsilon]\!]^{-1}\mathbf{K}_x + \mathbf{I} \\ 0 & 0 & \mathbf{K}_y[\![\epsilon]\!]^{-1}\mathbf{K}_y - \mathbf{I} & -\mathbf{K}_y[\![\epsilon]\!]^{-1}\mathbf{K}_x \\ -\mathbf{K}_x\mathbf{K}_y & \mathbf{K}_x^2 - [\![\epsilon]\!] & 0 & 0 \\ -\mathbf{K}_y^2 + [\![\epsilon]\!] & \mathbf{K}_y\mathbf{K}_x & 0 & 0 \end{pmatrix}}_{\mathbf{M}} \begin{pmatrix}\mathbf{S}_x\\ \mathbf{S}_y\\ \mathbf{U}_x\\ \mathbf{U}_y\end{pmatrix} \tag{2.11}$$

Here, $\mathbf{K}_j$ is a $MN \times MN$ diagonal matrix of the scaled wave vector components $k_{j;m,n}/k_0$. $[\![\epsilon]\!]$ is a Block-Toeplitz-Toeplitz-Block (BTTB) matrix of the permittivity coefficients $\epsilon_{m,n}$. A BTTB-matrix is basically a second order Toeplitz matrix that holds a four-dimensional dataset $[\![\epsilon]\!]_{m,n,q,r} = \epsilon_{m-n,q-r}$ in a conventional (2D) matrix form as depicted in Fig. 2.2.

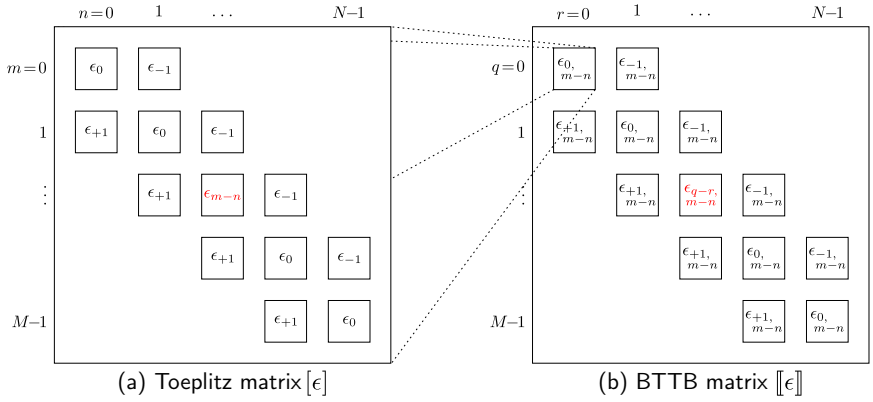


Figure 2.2: Illustration of a (a) Toeplitz matrix with $[\epsilon]_{m,n} = \epsilon_{m-n}$ and (b) a BTTB matrix with $[\![\epsilon]\!]_{m,n,q,r} = [\![\epsilon]\!]_{m\cdot N+q,n\cdot M+r} = \epsilon_{m-n,q-r}$

In Fourier space, the product of any two periodic space functions, such as $\mathbf{E}(x,y)$ and $\epsilon(x,y)$ can be expressed as a convolution of all its discrete Fourier coefficients. Depending on the dimension of the two functions, a sum notation given by the Laurent factorization allows a simple matrix-vector form by the use of a Toeplitz or BTTB matrix:

$$\epsilon(x,y)\cdot E(x,y) = D(x,y)/\epsilon_0 \tag{2.12a}$$

$$\xrightarrow{\mathscr{F}} \quad \epsilon_{m,n} * S_{m,n} = \sum_{q,r} \epsilon_{q-m,r-n} S_{m,n} \tag{2.12b}$$

$$= [\![\epsilon]\!]\cdot\mathbf{S} \tag{2.12c}$$

Since sums and matrices need to be finite in numerical calculations, mode truncation is inevitable but can lead to approximations for signals with high bandwidth.

For the sake of better readability the z-dependence in Eq. 2.12 and in the following equations is not written out until the solution of the eigenvalue problem provides a concrete dependency. Also, as mentioned earlier, the material parameters of the grating are all assumed to be of pure isotropic nature. Anisotropic behavior would be treated with an additional tensor degree[74], which would cause an unnecessary complexity here.

Relation 2.11 can also be written in a short form as

$$\frac{\partial}{\partial z}\begin{pmatrix}\mathbf{S}_\perp\\ \mathbf{U}_\perp\end{pmatrix} = \mathbf{M}\begin{pmatrix}\mathbf{S}_\perp\\ \mathbf{U}_\perp\end{pmatrix} \tag{2.13}$$

with $\mathbf{S}_\perp$ and $\mathbf{U}_\perp$ each denoting a stacked, $2MN$-sized vector of the Fourier coefficients and an anti-diagonal block matrix $\mathbf{M}$.

By identifying the block elements $\mathbf{M}_1$ and $\mathbf{M}_2$ of matrix $\mathbf{M}$ as

$$\frac{\partial}{\partial z}\begin{pmatrix}\mathbf{S}_\perp\\ \mathbf{U}_\perp\end{pmatrix} = \begin{pmatrix}0 & \mathbf{M}_1\\ \mathbf{M}_2 & 0\end{pmatrix}\begin{pmatrix}\mathbf{S}_\perp\\ \mathbf{U}_\perp\end{pmatrix} \tag{2.14}$$

with

$$\mathbf{M}_1 := ik_0\left[\begin{pmatrix}\mathbf{K}_x[\![\epsilon]\!]^{-1}\mathbf{K}_y & -\mathbf{K}_x[\![\epsilon]\!]^{-1}\mathbf{K}_x\\ \mathbf{K}_y[\![\epsilon]\!]^{-1}\mathbf{K}_y & -\mathbf{K}_y[\![\epsilon]\!]^{-1}\mathbf{K}_x\end{pmatrix} + \begin{pmatrix}0 & \mathbf{I}\\ -\mathbf{I} & 0\end{pmatrix}\right] \tag{2.15a}$$

$$\mathbf{M}_2 := ik_0\left[\begin{pmatrix}-\mathbf{K}_x\mathbf{K}_y & \mathbf{K}_x^2\\ -\mathbf{K}_y^2 & \mathbf{K}_y\mathbf{K}_x\end{pmatrix} + \begin{pmatrix}0 & -[\![\epsilon]\!]\\ [\![\epsilon]\!] & 0\end{pmatrix}\right], \tag{2.15b}$$

a relation between $\mathbf{S}_\perp$ and the second order derivative of $\mathbf{S}_\perp$ with respect to z can be found[4]:

$$\begin{aligned}\partial\mathbf{S}_\perp/\partial z &= \mathbf{M}_1\cdot\mathbf{U}_\perp\\ \partial^2\mathbf{S}_\perp/\partial z^2 &= \mathbf{M}_1\cdot\partial\mathbf{U}_\perp/\partial z = \mathbf{M}_1\mathbf{M}_2\cdot\mathbf{S}_\perp\\ \Rightarrow \partial^2\mathbf{S}_\perp/\partial z^2 - \mathbf{\Omega}_E\cdot\mathbf{S}_\perp &= 0 \qquad \text{with } \mathbf{\Omega}_E := \mathbf{M}_1\mathbf{M}_2\end{aligned} \tag{2.16}$$

Equation 2.16 shows a system of second order ordinary differential equations with constant coefficients that has the same form of the Helmholtz equation (cf. Eq. 1.39) and can similarly

[4] An equivalent relation for the magnetic field is derived in Sec. A.5

be transformed into an eigenvalue problem: if $\mathbf{V}$ is the matrix of eigenvectors of $\mathbf{\Omega}_E$ and if $\boldsymbol{\lambda}$ is the vector of eigenvalues, then the Fourier coefficients of the electric field can be written as

$$\mathbf{S}_\perp = \mathbf{V}\mathbf{A} \tag{2.17}$$

while using the exponential approach (cf. 1.41)

$$A_q(z) = t_{\perp;q} e^{-k_0\sqrt{\lambda_q}z} + r_{\perp;q} e^{k_0\sqrt{\lambda_q}(z-d)}. \tag{2.18}$$

Thus, in Eq. 2.17 an expression is found that describes the electric field inside the grating region without the dependence of a magnetic field and without derivatives of any order.

In Eq. 2.18 q is an index ranging from 0 to $2MN-1$. Each coefficient $t_{\perp;q}$ and $r_{\perp;q}$, which is yet to be determined, can be identified as the amplitude of a fundamental mode or a plane wave, which solves the diffraction problem and propagates undisturbed in the grating region in forward or backward direction while experiencing an effective index $n_q = i\sqrt{\lambda_q}/s_{z;q}$ (cf. Fig. 2.3). For positive real values of $\sqrt{\lambda_q}$ the opposing plane waves become evanescent and thus decay in distance to the borders (cf. red curves in Fig. 2.3).

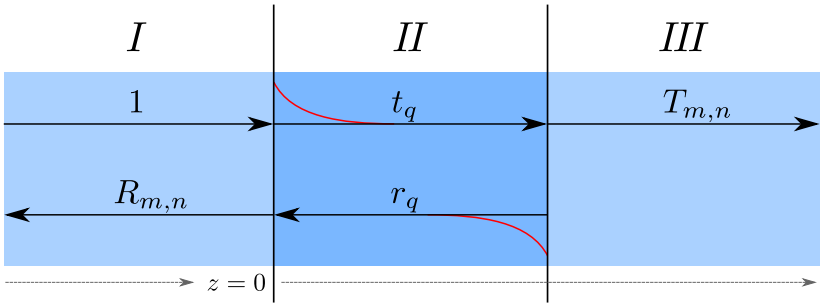


Figure 2.3: Forward and backward propagating plane waves in different regions of the simulation area including evanescent waves in red

The exponential approach in Eq. 2.18 can easily be shown to be a solution of the eigenvalue problem[126]:

$$\begin{aligned}
\partial^2\mathbf{S}_\perp/\partial z^2 - \mathbf{\Omega}_E \cdot \mathbf{S}_\perp &= \partial^2\mathbf{V}\mathbf{A}/\partial z^2 - \mathbf{\Omega}_E\left(\mathbf{V}\mathbf{A}\right) \\
&= \mathbf{V}\partial^2\mathbf{A}/\partial z^2 - \mathbf{\Omega}_E\mathbf{V}\mathbf{A} \\
&= \mathbf{V}\mathbf{\Lambda}\mathbf{A} - \mathbf{\Omega}_E\mathbf{V}\mathbf{A} \\
&= \left(\mathbf{V}\mathbf{\Lambda} - \mathbf{\Omega}_E\mathbf{V}\right)\mathbf{A} \\
&= 0
\end{aligned}$$

With the following three definitions made

$$\Gamma_q := \sqrt{\lambda_q} \tag{2.19a}$$
$$\mathbf{W} := \mathbf{M_1^{-1}V\Gamma} \tag{2.19b}$$
$$B_q(z) := -t_{\perp;q}e^{-k_0\sqrt{\lambda_q}z} + r_{\perp;q}e^{k_0\sqrt{\lambda_q}(z-d)}, \tag{2.19c}$$

the Fourier coefficients of the magnetic field can be derived from Eq. 2.20, which saves the effort of solving a second eigenvalue problem:

$$\begin{aligned}\mathbf{U}_\perp &= \mathbf{M}_1^{-1}\partial\mathbf{S}_\perp/\partial z \\ &= \mathbf{M_1^{-1}V\Gamma B} \\ &= \mathbf{WB}\end{aligned} \tag{2.20}$$

Together, the Fourier coefficients of the electric and magnetic field components inside the grating can now be described as two separated equations

$$\mathbf{S}_\perp = \mathbf{VA} \tag{2.21a}$$
$$\mathbf{U}_\perp = \mathbf{WB} \tag{2.21b}$$

or as a one joint equation

$$\begin{pmatrix}\mathbf{S}_\perp \\ \mathbf{U}_\perp\end{pmatrix} = \begin{pmatrix}\mathbf{VP}^+ & \mathbf{VP}^- \\ \mathbf{WP}^+ & -\mathbf{WP}^-\end{pmatrix}\begin{pmatrix}\mathbf{t}_\perp \\ \mathbf{r}_\perp\end{pmatrix} \tag{2.22}$$

with $P_q^+ := e^{-k_0\sqrt{\lambda_q}z}, P_q^- := e^{k_0\sqrt{\lambda_q}(z-d)}$.

Rewritten, this equation shows the relation between two different representations of the field inside the grating.

$$\underbrace{\begin{pmatrix}\mathbf{S}_\perp(z) \\ \mathbf{U}_\perp(z)\end{pmatrix}}_{\mathbf{F}(z)} = \underbrace{\begin{pmatrix}\mathbf{V} & \mathbf{V} \\ \mathbf{W} & -\mathbf{W}\end{pmatrix}}_{\mathbf{Q}} \underbrace{\begin{pmatrix}\mathbf{P}^+(z) & 0 \\ 0 & \mathbf{P}^-(z)\end{pmatrix}}_{\mathbf{P}(z)} \underbrace{\begin{pmatrix}\mathbf{t}_\perp \\ \mathbf{r}_\perp\end{pmatrix}}_{\mathbf{\Psi}} \tag{2.23}$$

Fourier modes — Conversion + Coupling — Propagation — Fundamental modes

On the left side of the equation $\mathbf{F}(z)$ holds the coefficients of the lateral Fourier expansion of the electric and the magnetic field. On the right side $\mathbf{\Psi}$ contains the coefficients of the

field's eigenmodes in the grating. The propagation matrix $\mathbf{P}(z)$ propagates plane waves with amplitude $\mathbf{t}_{\perp,q}$ and $\mathbf{r}_{\perp,q}$ in forward and backward direction along z. And the matrix of eigenvectors $\mathbf{Q}$ transforms the eigenmodes into Fourier modes and thus describes their coupling in the grating region.

Boundary conditions

The remaining unknown quantities are the expansion coefficients inside the grating ($\mathbf{t}_\perp$ and $\mathbf{r}_\perp$, cf. Eq. 2.23) and outside ($\mathbf{T}_\perp$ and $\mathbf{R}_\perp$, cf. Eq. 2.6). They can be determined by taking into account Maxwell's boundary conditions of the tangential field components at the interfaces (cf. chapter 1.1.7) or more precisely by matching the fields using the Fourier coefficients as a common basis.

Eq. 2.23, which gives an expression for the complete electromagnetic field in the grating region, can in fact also be defined for the outer regions. With index $l = I, II, III$ indicating a particular region, Eq. 2.24 describes the relation between Fourier- and eigenmodes and holds in the entire simulation domain:

$$\mathbf{F}_l(z) = \mathbf{Q}_l \cdot \mathbf{P}_l(z) \cdot \mathbf{\Psi}_l \tag{2.24}$$

Since in the homogeneous outer regions I and III the Fourier expansion of the field already provides the fundamental modes, the expansion coefficients in these regions can be identified by some partially known quantities of Eq. 2.6:

$$\mathbf{t}_I := \mathbf{L} \tag{2.25a}$$
$$\mathbf{r}_I := \mathbf{R} \tag{2.25b}$$
$$\mathbf{t}_{III} := \mathbf{T} \tag{2.25c}$$
$$\mathbf{r}_{III} := 0. \tag{2.25d}$$

Eq. 2.25a shows a vector $\mathbf{L} := (\mathbf{L}_x\ \mathbf{L}_y\ \mathbf{L}_z)^T$ of size $2NM \times 1$, which holds the incident field vector. It is defined by the components of the only (0th order) Fourier mode of the incident plane wave from Eq. 2.6b:

$$L_{j;m,n} = E_{1;j}\ \delta_{m,m_0}\delta_{n,n_0} \tag{2.26a}$$

$$\Rightarrow \mathbf{L} = \begin{pmatrix} [cos(\psi)cos(\theta)cos(\phi) - sin(\psi)sin(\phi)]\ \vec{\delta}_{m,m_0;n,n_0} \\ [cos(\psi)cos(\theta)sin(\phi) + sin(\psi)cos(\phi)]\ \vec{\delta}_{m,m_0;n,n_0} \\ -cos(\psi)sin(\theta)\vec{\delta}_{m,m_0;n,n_0} \end{pmatrix}. \tag{2.26b}$$

$\delta_{a,b}$ is the Kronecker delta, which is zero for all integer values of a except for the value $a = b$, where it defined as 1. The special notation $\vec{\delta}_{m,m_0}$ in Eq. 2.26b indicates a vector of length NM with a single non-zero entry at the exact center index $m_0 = (2M_0 + 1) \cdot N_0 + M_0$.

According to Eq. 2.24 the tangential Fourier coefficients of the electric and the magnetic field at the outer boundaries can then be described as

$$\mathbf{F}_I\ (z_I\ = -0) = \mathbf{Q}_I\ \cdot \mathbf{P}_I\ \cdot \mathbf{\Psi}_I\ = \begin{pmatrix} \mathbf{I} & \mathbf{I} \\ \mathbf{C} & -\mathbf{C} \end{pmatrix}_I \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{L}_\perp \\ \mathbf{R}_\perp \end{pmatrix} \tag{2.27a}$$

$$\mathbf{F}_{III}(z_{III} = +0) = \mathbf{Q}_{III} \cdot \mathbf{P}_{III} \cdot \mathbf{\Psi}_{III} = \begin{pmatrix} \mathbf{I} & \mathbf{I} \\ \mathbf{C} & -\mathbf{C} \end{pmatrix}_{III} \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{T}_\perp \\ 0 \end{pmatrix}. \tag{2.27b}$$

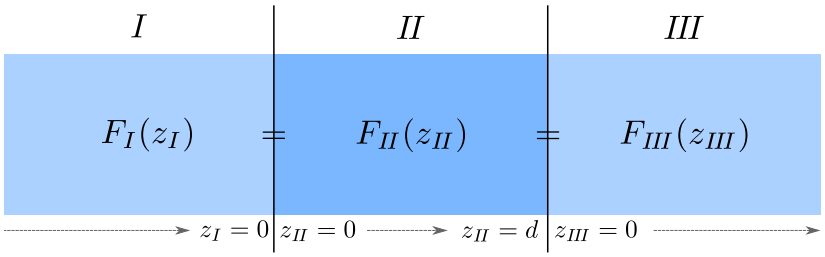


Figure 2.4: Electromagnetic fields with locally defined z-positions and matching boundary conditions at both grating interfaces

With the definition of a local z position according to Fig. 2.4, no propagation is needed at the interfaces and thus it is $\mathbf{P}_I = \mathbf{P}_{III} = \mathbf{I}$ with $\mathbf{I}$ as the identity matrix. In the homogeneous medium of region I and III the field is already formed by Fourier modes, which are equal to the eigenmodes. Therefore, the coupling and conversion matrix $\mathbf{Q}$ simplifies to a matrix that only describes the conversion between electric and magnetic coefficients.

The exact definition of the 2×2 block matrix $\mathbf{C}$ follows from the requirement $\mathbf{U}_\perp = \mathbf{C} \cdot \mathbf{S}_\perp$ (cf. Appendix A.4) and leads to

$$\mathbf{C}_l := \frac{1}{\mu k_0 Z_0} \begin{bmatrix} -\operatorname{diag}_q\left(\dfrac{k_{x;q} k_{y;q}}{k_{z;l;q}}\right) & -\operatorname{diag}_q\left(\dfrac{k_{y;q}^2 + k_{z;l;q}^2}{k_{z;l;q}}\right) \\ \operatorname{diag}_q\left(\dfrac{k_{x;q}^2 + k_{z;l;q}^2}{k_{z;l;q}}\right) & \operatorname{diag}_q\left(\dfrac{k_{x;q} k_{y;q}}{k_{z;l;q}}\right) \end{bmatrix}. \tag{2.28}$$

The operator $\operatorname{diag}_q(v_q)$ converts all elements v_q with $q = 0 \,..\, \operatorname{count}(\mathbf{v}) - 1$ of a vector $\mathbf{v}$ into a diagonal matrix with $\operatorname{count}(\mathbf{v}) \cdot \operatorname{count}(\mathbf{v})$ elements[5]. For the matrix $\mathbf{C}$ the index l refers to one of the outer regions $l = I$ or $l = III$.

[5]The operator $\operatorname{count}(\mathbf{v})$ gives the number of elements of the vector $\mathbf{v}$.

The Fourier coefficients at the surface boundaries can also be defined from the inside of the grating by

$$\mathbf{F}_{II}(0) = \mathbf{Q}_{II}\mathbf{P}_{II}(0)\mathbf{\Psi}_{II} \qquad \text{and} \tag{2.29a}$$
$$\mathbf{F}_{II}(d) = \mathbf{Q}_{II}\mathbf{P}_{II}(d)\mathbf{\Psi}_{II}. \tag{2.29b}$$

Thus, by following Maxwell's continuity conditions as depicted in Fig. 2.4, the tangential field components at the interface boundaries can now be set equal, leading to

$$\mathbf{F}_{I}(0) \overset{!}{=} \mathbf{F}_{II}(0) \qquad \text{and} \tag{2.30a}$$
$$\mathbf{F}_{III}(0) \overset{!}{=} \mathbf{F}_{II}(d) \tag{2.30b}$$

$$\Leftrightarrow$$

$$\begin{pmatrix} \mathbf{I} & \mathbf{I} \\ \mathbf{C} & -\mathbf{C} \end{pmatrix}_{I} \begin{pmatrix} \mathbf{L}_{\perp} \\ \mathbf{R}_{\perp} \end{pmatrix} \overset{!}{=} \begin{pmatrix} \mathbf{V} & \mathbf{V} \\ \mathbf{W} & -\mathbf{W} \end{pmatrix}_{II} \begin{pmatrix} \mathbf{I} & 0 \\ 0 & \mathbf{X} \end{pmatrix}_{II} \begin{pmatrix} \mathbf{t}_{\perp} \\ \mathbf{r}_{\perp} \end{pmatrix}_{II} \tag{2.31a}$$
$$\begin{pmatrix} \mathbf{I} & \mathbf{I} \\ \mathbf{C} & -\mathbf{C} \end{pmatrix}_{III} \begin{pmatrix} \mathbf{T}_{\perp} \\ 0 \end{pmatrix} \overset{!}{=} \begin{pmatrix} \mathbf{V} & \mathbf{V} \\ \mathbf{W} & -\mathbf{W} \end{pmatrix}_{II} \begin{pmatrix} \mathbf{X} & 0 \\ 0 & \mathbf{I} \end{pmatrix}_{II} \begin{pmatrix} \mathbf{t}_{\perp} \\ \mathbf{r}_{\perp} \end{pmatrix}_{II} \tag{2.31b}$$

with $X_{q,q} = e^{-k_0\sqrt{\lambda_q}d}$.

Eq. 2.31 shows a system of four matrix equations with four unknown vectors $\mathbf{T}_{\perp}$, $\mathbf{R}_{\perp}$, $\mathbf{t}_{\perp}$ and $\mathbf{r}_{\perp}$. In order to determine the reflection and transmission coefficients, the system can now be solved by first eliminating $\mathbf{T}_{\perp}$, $\mathbf{R}_{\perp}$ and then substituting $\mathbf{t}_{\perp}$, $\mathbf{r}_{\perp}$:

$$\begin{pmatrix} \mathbf{t}_{\perp} \\ \mathbf{r}_{\perp} \end{pmatrix} = \begin{pmatrix} \mathbf{C}_I\mathbf{V} + \mathbf{W} & [\mathbf{C}_I\mathbf{V} - \mathbf{W}]\,\mathbf{X} \\ -[\mathbf{C}_{III}\mathbf{V} - \mathbf{W}]\,\mathbf{X} & -[\mathbf{C}_{III}\mathbf{V} + \mathbf{W}] \end{pmatrix}^{-1} \begin{pmatrix} 2\mathbf{C}_I\mathbf{L}_{\perp} \\ 0 \end{pmatrix} \tag{2.32}$$

$$\Downarrow$$

$$\mathbf{R}_{\perp} = \mathbf{V}\mathbf{t}_{\perp} + \mathbf{V}\mathbf{X}\mathbf{r}_{\perp} - \mathbf{L}_{\perp} \tag{2.33a}$$
$$\mathbf{T}_{\perp} = \mathbf{V}\mathbf{X}\mathbf{t}_{\perp} + \mathbf{V}\mathbf{r}_{\perp} \tag{2.33b}$$

These diffraction coefficients are the result of the standard RCWA algorithm.

Remark: The submatrix **C** in Eq. 2.28 is defined in a global Cartesian coordinate system with the base vectors $_x$, $_y$ and $_z$. This results in diffraction coefficients **R** and **T** being also defined in a global Cartesian coordinate system rather than local TE-/TM-systems of the individual diffraction order with base vectors $e_{N;m,n}$, $e^{I}_{T;m,n}$, $e^{III}_{T;m,n}$ (cf. App. A.6). This spares the need for local rotation matrices that are usually used to solve the conical problem[91]. As a result the concept of the conversion matrix **Q** is adopted to the outside region (cf. Eqs. 2.27), so that Maxwell's continuity condition at the grating interfaces (cf. Eqs. 2.31) can be satisfied in a way that is as straight forward and simple as in the classical (non-conical) diffraction problem[91]. This approach was also published in [6]. A conversion to the local TE-/TM-coordinate systems can also be reviewed in App. A.6.

2.5 Convergence problems

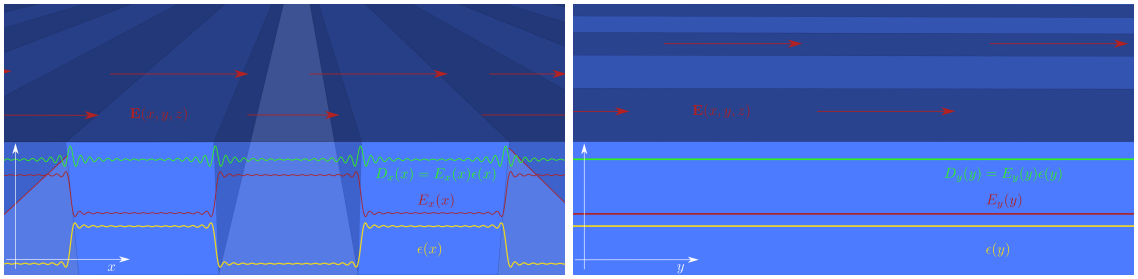
The convergence of a numerical algorithm describes how the error between the numerical solution and the exact solution changes over the number of iterations. By definition an algorithm has a good convergence behavior if the error is shrinking below a given limit in only a few iterations. If a lot of iterations are necessary to reach this limit, the convergence behavior is bad. And if the error is not shrinking at all, the algorithm is called divergent.
A necessary, even though not sufficient condition for the convergence to an exact solution, is any convergence at all. Since in most scenarios, in which the RCWA is used, the exact solution is unknown, the algorithm is repeated with increasing truncation order until the difference of two subsequent solution shrinks below a given limit.

As a Fourier modal method the RCWA has a fundamental issue in dealing high frequencies as they appear at the steep slopes of binary grating. This is because of the necessity of truncating the Fourier series to a finite limit in numerical calculations. Especially jump discontinuities, as they occur in a rectangular function, contain an infinite number of frequencies and a truncated Fourier approximation results in the appearance of over- and undershoots - well known as the Gibbs phenomenon. Even with an increasing number of modes to model the signal, the effect cannot be eliminated. However, the use of suitable window functions, as they are broadly used in signal processing[45], can - at the cost of spatial resolution - reduce the effect. In optics this is technique is known as apodization and discussed in Sec. 2.6.

In certain cases, though, the effect of those over- and undershoots can also cause oscillations in field components that are supposed to be continuous and should not contain high frequencies. This lead to a significant decrease of the convergence speed.

2.5.1 One-dimensional gratings, the origin of the problem and its acknowledged solution

After the first publication of the RCWA for one-dimensional gratings by Moharam and Gaylord, it became evident, that the algorithm suffered bad convergence behavior in case of TM polarized incident light ($\psi = 0$ deg, cf. Sec. 2.3.3). This was caused by unphysical oscillations of the displacement field: In the TM case the electric field vector $\mathbf{E}(\mathbf{x})$ always points normal to the one-dimensional binary gratings grooves (cf. Fig. 2.5a) and develops a pairwise-concurrent jump discontinuity with the spacial permittivity $\epsilon(x)$. As a result of the truncated Fourier series expansion, the electric displacement field[6] $D_x(x) = \epsilon_0 \epsilon(x) E_x(x)$, which should be continuous in normal direction (cf. 1.1.7), show over- and undershoots due to the Gibbs phenomena of the constructing functions. This leads to bad convergence behavior.



(a) TM polarization (b) TE polarization

Figure 2.5: Difference between TM and TE polarization at a one-dimensional grating (modulation along x): (a) TM polarization causes unphysical oscillations at the grating slopes resulting in a bad convergence behavior.
(b) TE polarization is harmless since no jump discontinuities occur.

A solution was first found empirically: In 1996, Lalanne and Morris[68] and also Granet and Guizal[41], independently demonstrated significantly improved convergence for the TM case by replacing the Toeplitz matrices $[\epsilon]$ by $[1/\epsilon]^{-1}$ in the product with the electric field components $\mathbf{S}_x$ normal to grating slopes. Shortly thereafter, this idea was given a theoretical basis by Li[77] and therein referred to as the inverse rule. Li also embedded the rule into a greater set of rules (later referred to as Li's factorization rules), which give guidance to the correct use of the inverse rule in different situations. A year later he extended the factorization rules to the special case of two-dimensional crossed gratings[76]. A more general approach to apply the inverse rule in two-dimensional gratings uses a normal vector field. It was first proposed for the differential method by Popov and Nevière[107] and later reformulated for the RCWA by Schuster[112, 113] (cf. Sec. 2.5.5). The latter approach is not pursued further in this work. However, it is worthwhile mentioning as it shows considerably improved performance in the case of curved grating structures compared to Li's formulation that relies on a zigzag approximation for such case.

[6] The following discussion focuses on normal incidence, TM polarization and one-dimensional gratings with grating slopes aligned to the y-axis. Thus only an x-component of the electric field exists and the subscript is left out.

2.5.2 Li's rules

In order to preserve the smoothness of the field across discontinuities of the permittivity, Li concluded three factorization rules[77], later known as the Li's rules, which define the appropriate usage of the inverse rule:

1. A product of two piecewise-smooth, bounded, periodic functions that have no concurrent jump discontinuities can be Fourier factorized by Laurent's rule.

$$\text{Example:} \qquad D(x) = \epsilon_0 \epsilon(x) E(x) \quad \xrightarrow{\mathscr{F}} \quad \mathbf{D} = \epsilon_0 \, [\epsilon] \, \mathbf{S} \tag{2.34}$$

2. A product of two piecewise-smooth, bounded, periodic functions that have only pairwise-complementary jump discontinuities cannot be Fourier factorized by Laurent's rule, but in most cases it can be Fourier factorized by the inverse rule.

$$\text{Example:} \qquad D(x) = \epsilon_0 \epsilon(x) E(x) \quad \xrightarrow{\mathscr{F}} \quad \mathbf{D} = \epsilon_0 \, [1/\epsilon]^{-1} \, \mathbf{S} \tag{2.35}$$

3. A product of two piecewise-smooth, bounded, periodic function that have concurrent but not complementary jump discontinuities can be Fourier factorized by neither Laurent's rule nor the inverse rule.

In his first rule, Li claims that for non-concurrent jump discontinuities, the Laurent's factorization rule applies as usual. However, it should be noted that the Gibbs phenomenon that occurs at sharp edges, still remains present as a fundamental problem of Fourier expansion.

In case the two functions $\epsilon(x)$ and $E(x)$ do have pairwise-concurrent jump discontinuities, Li first derives the Fourier coefficients of $D(x) := \epsilon_0 \epsilon(x) E(x)$ from the convolution of $\mathscr{F}\{\epsilon_0 \epsilon(x) \cdot E(x)\} = \epsilon_0 \epsilon_m * S_m$ according to Laurent's rule (cf. Eq. 2.36a). Then, he defines the Fourier expansion of $D(x)$ with truncated Fourier coefficients $D_n^{(M_0)}$ and the same summation limits (cf. Eq. 2.36b). And finally he defines another Fourier expansion of $D(x)$ with exact Fourier coefficients D_n (cf. Eq. 2.36c):

$$D_n^{(M_0)} := \epsilon_0 \sum_{m=-M_0}^{M_0} \epsilon_{n-m} S_m \tag{2.36a}$$

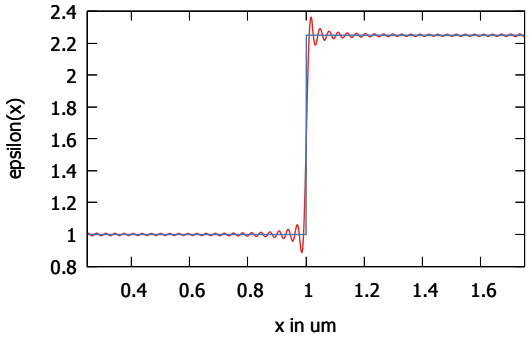
$$D^{(M_0)}(x) := \sum_{n=-M_0}^{M_0} D_n^{(M_0)} \, e^{inx} \tag{2.36b}$$

$$D_{M_0}(x) := \sum_{n=-M_0}^{M_0} D_n \, e^{inx} \tag{2.36c}$$

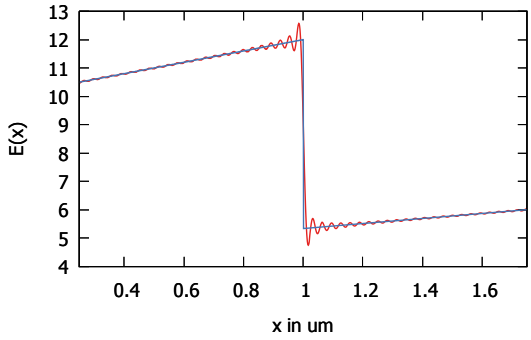
Subscript M or superscript M enclosed in parentheses denote the symmetrically truncated partial sums. Although a reader would expect $D^{(M_0)}(x) = D_{M_0}(x)$ as M_0 is tending to infinity, Li showed that in general for concurrent jumps, the difference between the two does not vanish everywhere, but tends to a fixed limit of $\pi^2/4$ at the position of the jump[77]. However, by replacing the Toeplitz matrix according to the inverse rule

$$\hat{D}_n^{(M_0)} := \epsilon_0 \sum_{m=-M_0}^{M_0} [1/\epsilon]_{n,m}^{-1} \; S_m, \tag{2.37}$$

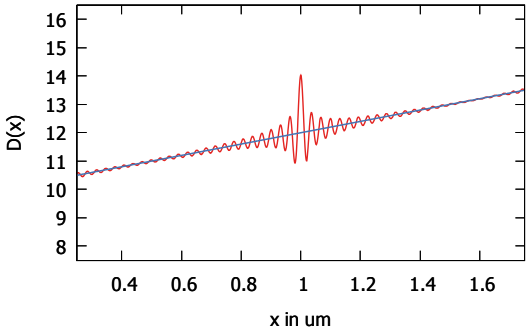
the over- and undershoots of the concurrent jump discontinuities compensate for each other and disappear in the course of the convolution operation. This is illustrated in Figs. 2.6a to 2.6d.



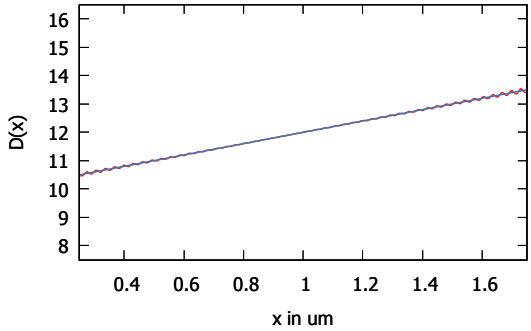
(a) $\epsilon(x)$ in blue, $\epsilon^{(M_0)}(x)$ in red with Gibbs phenomenon



(b) $E(x)$ in blue, $E^{(M_0)}(x)$ in red with Gibbs phenomenon



(c) $D(x)$ in blue, $D^{(M_0)}(x)$ in red with unphysical oscillations by Laurent's rule



(d) $D(x)$ in blue, $\hat{D}^{(M_0)}(x)$ in red with restored smoothness by the inverse rule

Figure 2.6: Exemplary comparison between Laurent's rule and the inverse rule for the use of concurrent jump discontinuities in Fourier reconstruction

Consistency of the inverse rule can easily be derived:

$$D(x) = \epsilon_0 \epsilon(x) E(x) \quad \xrightarrow{\mathscr{F}} \quad \mathbf{D} = \epsilon_0 \, [\epsilon] \, \mathbf{S} \tag{2.38a}$$

$$\Rightarrow \quad 1/\epsilon_0 1/\epsilon(x) D(x) = E(x) \quad \xrightarrow{\mathscr{F}} \quad 1/\epsilon_0 \, [1/\epsilon] \, \mathbf{D} = \mathbf{S} \tag{2.38b}$$

$$\Rightarrow \quad \mathbf{D} = \epsilon_0 \, [1/\epsilon]^{-1} \, \mathbf{S} \tag{2.38c}$$

This shows the difference between an untruncated calculation on paper and a truncated numerical calculation. In the first case all vectors and matrices in Eqs. 2.38a - 2.38c have infinite size and the above transformations are exact, but also irrelevant, because the two replacing matrices are identical. In the second case, the truncation selects frequencies in the same interval but of the two different spectra of $\epsilon(x)$ and $1/\epsilon(x)$. Apparently this leads to a slight difference of $[\epsilon]$ and $[1/\epsilon]^{-1}$ that grows with decreasing truncation order. However, as mentioned above, while $[\epsilon]$ and $\mathbf{S}$ have concurrent over- and undershoots, which add up and cause the unphysical oscillations of $D(x)$ in the first place, $[1/\epsilon]^{-1}$ and $\mathbf{S}$ have compensating over- and undershoots and thus resolve the convergence problem of the overall algorithm for one-dimensional gratings.

2.5.3 Example

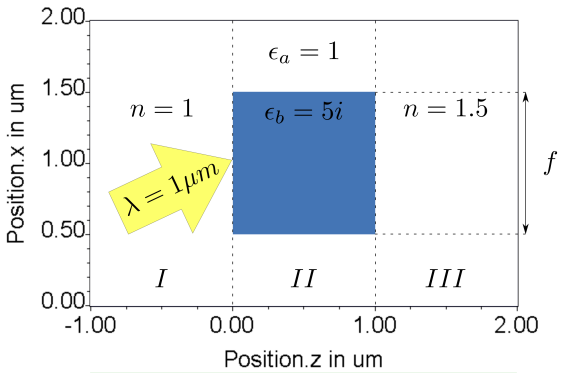
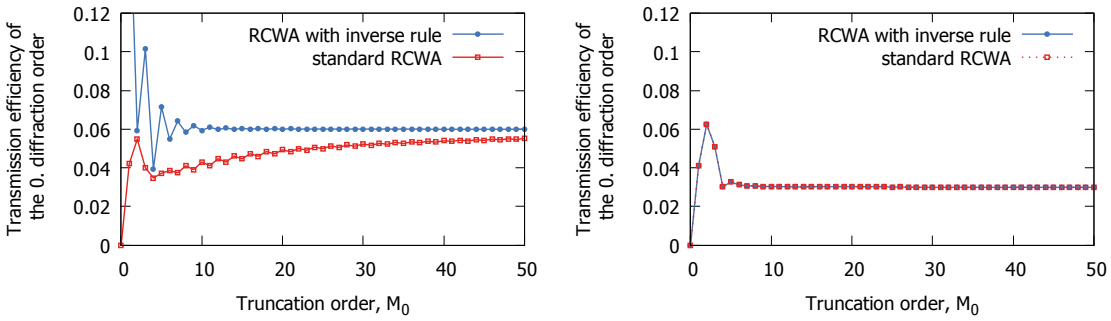


Figure 2.7: Sketch of a Ronchi metal grating

In order to demonstrate the influence of the inverse rule on the convergence behavior of the RCWA, a metal grating as depicted in Fig. 2.7 is analyzed, since those are known to cause bad convergence in the standard approach if illuminated with TM polarized light[76]. The grating is excited by only a single plane wave with wavelength $\lambda = 1\,\mu m$ at an incidence angle of $\theta = 30$ deg. Superstrate and substrate are chosen to have a refractive index of $n_I = 1$ and $n_{III} = 1.5$ respectively, while the grating is a Ronchi grating with a duty cycle of $f = 0.5$ at a period of $P_x = 2\lambda$. This causes the grating to alternate its refractive index along the x-axis between $\epsilon_{II,a} = 1$ and $\epsilon_{II,b} = 0 + 5i$. The grating's thickness is $d = \lambda$.

Fig. 2.8a clearly shows the superior convergence of the RCWA in case of TM polarized light when the inverse rule is applied. The results specifically refer to a one-dimensional grating that is periodic in the direction of x. Since the same grating is constant is the direction of y, there is no need for the inverse rule in the TE case and the modified RCWA naturally shows its unaltered good convergence behavior in Fig. 2.8b.



(a) TM polarization

(b) TE polarization

Figure 2.8: Convergence speed of the standard RCWA and the RCWA with inverse rule applied, when analyzing a one-dimensional grating with periodicity in the direction of x.

2.5.4 Two-dimensional gratings

For two-dimensional gratings the situation is more complex, because the field component normal to structural boundaries can vary dependent on the position inside a layer. Therefore, the convergence problem is not confined to the TM case anymore.

In 1996 Li proposed a reformulation of his factorization rules for two-dimensional gratings[76]. It refers to the special case of crossed grating boundaries that are aligned along two independent basis vectors in the x-y-plane. Although curved geometries can also be treated with this approach, they first need to be approximated with zigzag patterns.

For his approach Li introduces a new notation for partly Fourier transformed coefficients:

$$\lceil\epsilon\rceil_{m,n}(y) = \frac{1}{P_x}\int_0^{P_x} \epsilon(x,y)e^{-2\pi i\frac{m-n}{P_x}x}dx \tag{2.39a}$$

$$\lfloor\epsilon\rfloor_{m,n}(x) = \frac{1}{P_y}\int_0^{P_y} \epsilon(x,y)e^{-2\pi i\frac{m-n}{P_y}y}dy \tag{2.39b}$$

and

$$\lfloor\lceil\epsilon\rceil\rfloor_{m,n;q,r} = \left\lfloor\left[\frac{1}{\epsilon}\right]^{-1}_{m,q}\right\rfloor_{n,r} = \frac{1}{P_y}\int_0^{P_y}\{\lceil 1/\epsilon\rceil^{-1}\}_{m,q}(y)e^{-2\pi i\frac{n-r}{P_y}y}dy \tag{2.40a}$$

$$\lceil\lfloor\epsilon\rfloor\rceil_{m,n;q,r} = \left\lceil\left\lfloor\frac{1}{\epsilon}\right\rfloor^{-1}_{n,r}\right\rceil_{m,q} = \frac{1}{P_x}\int_0^{P_x}\{\lfloor 1/\epsilon\rfloor^{-1}\}_{m,q}(x)e^{-2\pi i\frac{m-q}{P_x}x}dx \tag{2.40b}$$

First, in Eq. 2.39, a Fourier expansion is applied for only one dimension and the resulting coefficients are assembled in Toeplitz form along the same direction. In 2.40 Li applies the inverse rule. A second Fourier expansion is applied to the complementing dimension along with another Toeplitz reordering, which finally leads to a BTTB matrix.

This way, $\epsilon(x,y)$ is fully transformed into Fourier space, but the inverse rule is only applied to one dimension. Thereby, discontinuities can be treated separately according to their orientation. To give a more specific example: The product $\lfloor\lceil\epsilon\rceil\rfloor\mathbf{S}_x$ would treat pair-wise concurrent jump discontinuities along the x-axis with the inverse rule to ensure smooth continuity of the displacement field while using Laurent's factorization rule for the transformation in y-direction.

In summary, for two-dimensional crossed gratings, Li's second factorization rule can be reformulated to the following:

If $\epsilon(x,y)$ and $E(x,y)$ are two piecewise-smooth, bounded functions with periodicity in x and y, that have only concurrent jump discontinuities along a single dimension, then their product can be Fourier factorized by applying the inverse rule to the discontinuous and Laurent's rule to the continuous dimension.

$$\text{Examples:} \quad D_x(x,y) = \epsilon_0 \epsilon(x,y) E_x(x,y) \quad \xrightarrow{\mathscr{F}} \quad \mathbf{D}_x = \epsilon_0 \lfloor\!\lceil \epsilon \rceil\!\rfloor \mathbf{S}_x \tag{2.41a}$$

$$D_y(x,y) = \epsilon_0 \epsilon(x,y) E_y(x,y) \quad \xrightarrow{\mathscr{F}} \quad \mathbf{D}_y = \epsilon_0 \lceil\!\lfloor \epsilon \rfloor\!\rceil \mathbf{S}_y \tag{2.41b}$$

Applying Li's rule to the standard RCWA, which was described at the beginning of this chapter, requires the replacement of those BTTB matrices, which form a product with the normal field components, according to the inverse rule. That way, in the context of equation

$$\frac{\partial}{\partial z} \begin{pmatrix} \mathbf{S}_\perp \\ \mathbf{U}_\perp \end{pmatrix} = \begin{pmatrix} 0 & \mathbf{M}_1 \\ \mathbf{M}_2 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{S}_\perp \\ \mathbf{U}_\perp \end{pmatrix} \tag{2.42}$$

matrix $\mathbf{M}_1$ and $\mathbf{M}_2$ transforms into

$$\mathbf{M}_1 := ik_0 \left[\begin{pmatrix} \mathbf{K}_x [\![\epsilon]\!]^{-1} \mathbf{K}_y & -\mathbf{K}_x [\![\epsilon]\!]^{-1} \mathbf{K}_x \\ \mathbf{K}_y [\![\epsilon]\!]^{-1} \mathbf{K}_y & -\mathbf{K}_y [\![\epsilon]\!]^{-1} \mathbf{K}_x \end{pmatrix} + \begin{pmatrix} 0 & \mathbf{I} \\ -\mathbf{I} & 0 \end{pmatrix} \right] \tag{2.43a}$$

$$\mathbf{M}_2 := ik_0 \left[\begin{pmatrix} -\mathbf{K}_x \mathbf{K}_y & \mathbf{K}_x^2 \\ -\mathbf{K}_y^2 & \mathbf{K}_y \mathbf{K}_x \end{pmatrix} + \begin{pmatrix} 0 & -\lceil\!\lfloor \epsilon \rfloor\!\rceil \\ \lfloor\!\lceil \epsilon \rceil\!\rfloor & 0 \end{pmatrix} \right]. \tag{2.43b}$$

Li's (factorization) rules are widely used, but require all lateral grating slopes to be aligned to two independent base vectors. Curved structures or other geometries always have to be approximated by zigzag patterns as shown in Fig. 2.9.

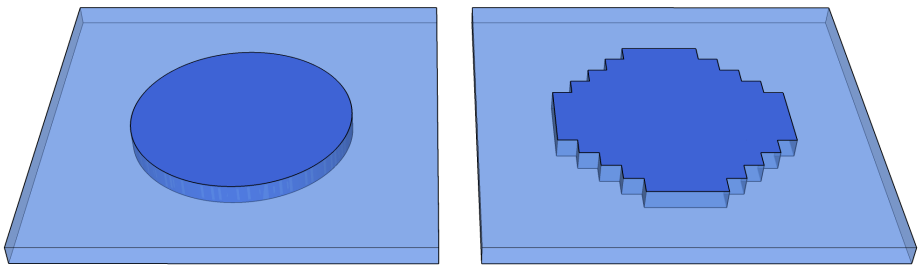


Figure 2.9: Zigzag approximation of a curved grating slope

2.5.5 Normal vector fields

A more general approach was first suggested 2001 by Popov und Neviere[107] for the integral method and adapted to the RCWA by Schuster[112, 113] in 2007. The idea is to separate the local electric field into its normal and tangential components with respect to the grating boundaries at every position in the layer.

If $\mathbf{E}$ is the electric field inside a grating and $\mathbf{N}$ is a vector field normal to the grating boundaries, then the electric field can be projected on to the respective normal and tangential vectors by

$$\mathbf{E}_N = \mathbf{N}\,(\mathbf{N}\cdot\mathbf{E}) = (\mathbf{N}\otimes\mathbf{N})\,\mathbf{E} \tag{2.44a}$$

$$\mathbf{E}_T = \mathbf{E} - \mathbf{E}_N. \tag{2.44b}$$

The symbol $\otimes$ denotes the outer product of two vectors. After separation, the field components are Fourier transformed and the appropriate factorisation rules are applied individually for the normal and tangential components.

$$\mathbf{D}_N = \left[\!\left[\frac{1}{\epsilon}\right]\!\right]^{-1} [\![\mathbf{N}\otimes\mathbf{N}]\!]\mathbf{E} \tag{2.45a}$$

$$\mathbf{D}_T = [\![\epsilon]\!]\mathbf{E} - \left[\!\left[\frac{\mathbf{1}}{\epsilon}\right]\!\right]^{-1} [\![\mathbf{N}\otimes\mathbf{N}]\!]\mathbf{E}. \tag{2.45b}$$

This way, the inverse rule can always be applied correctly for any arbitrary geometry throughout the grating region. However, the boundaries' normal vectors are only well defined at the boundaries themselves. An extrapolation to a complete normal vector field is generally not unique (cf. Fig. 2.10) and requires sophisticated methods, which are still topic of ongoing research[40].

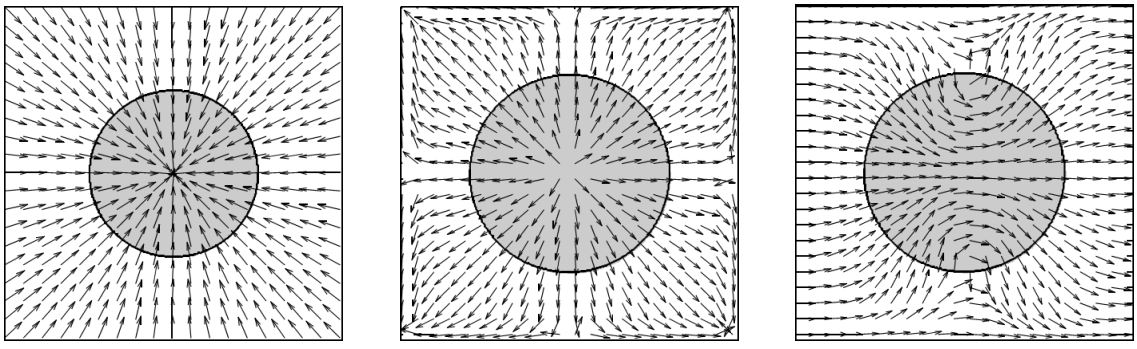


Figure 2.10: Normal vector field of a circular structure calculated with three different algorithms by Schuster. Source: T. Schuster. "Simulation von Lichtbeugung an Kreuzgitter-Strukturen und deren Anwendung in der Scatterometrie". PhD thesis. Institut für Technische Optik der Universitaet Stuttgart, 2010.

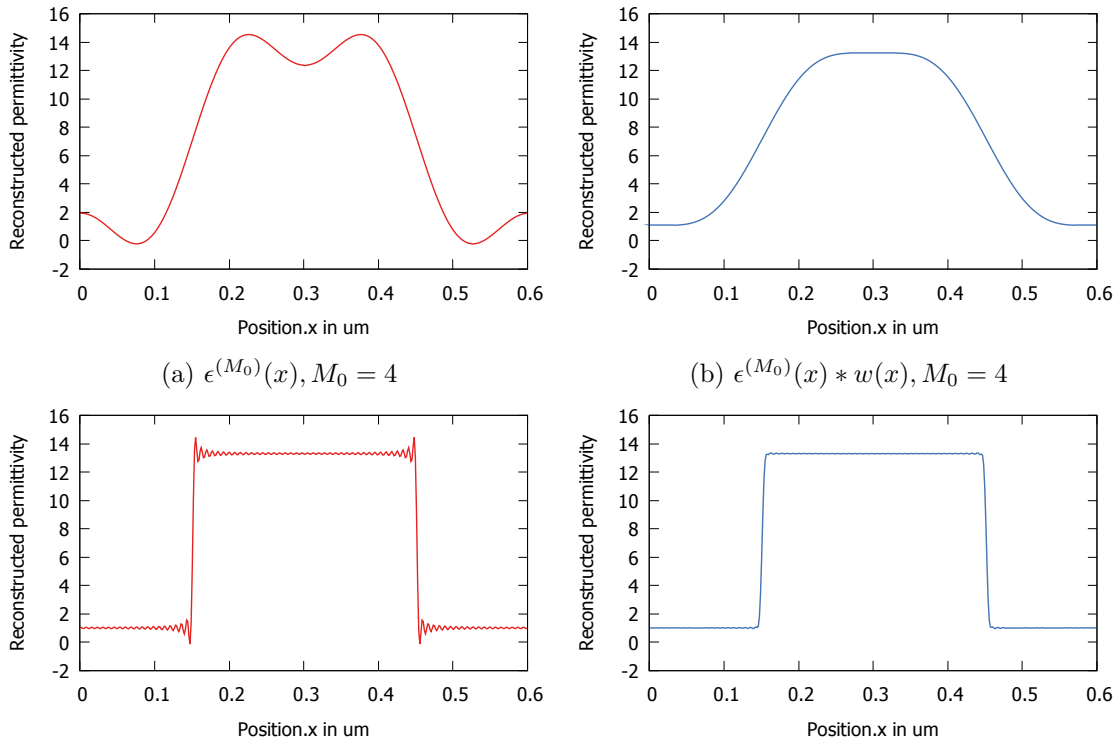
2.6 Apodization

In signal processing, the use of window functions, also known as apodization functions, are a common technique[45] to reduce the over- and undershoots of the Gibb's phenomena that occur when non-bandwidth limited functions are Fourier expanded. Thus, it seems reasonable to apply apodization to diffraction gratings, since it would target the root of the convergence problem in the RCWA. The idea is to multiply a window function with the signal (in this case the permittivity function) in the frequency domain and thereby fade out higher orders. In position space this corresponds with the smoothing of any sharp edges in the grating's structure.

A suitable apodization function is given by the folloing Gaussian function

$$a_m = e^{-2\left(\frac{m}{M_0}\right)^2}, \tag{2.46}$$

where m is the mode index. Figs. 2.11b and 2.11d show the smoothing effect of the apodization function for two different truncation orders: $M_0 = 4 \Rightarrow M = 2M_0 + 1 = 9$ and $M_0 = 80 \Rightarrow M = 161$



(a) $\epsilon^{(M_0)}(x), M_0 = 4$

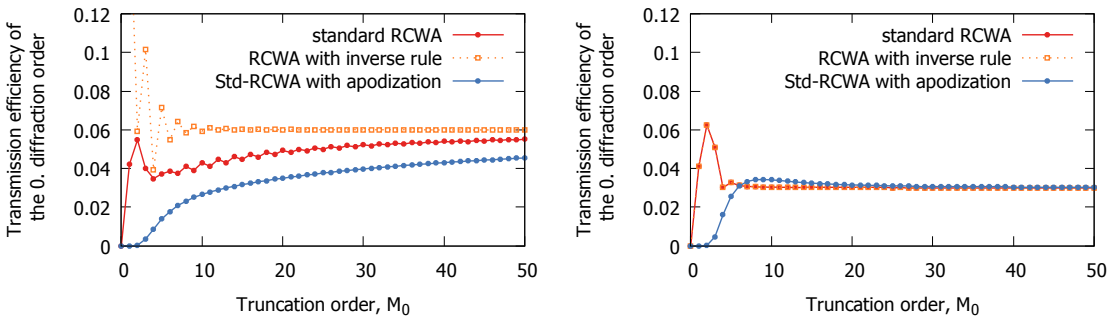
(b) $\epsilon^{(M_0)}(x) * w(x), M_0 = 4$

(c) $\epsilon^{(M_0)}(x), M_0 = 80$

(d) $\epsilon^{(M_0)}(x) * w(x), M_0 = 80$

Figure 2.11: Example for apodization at different truncation orders

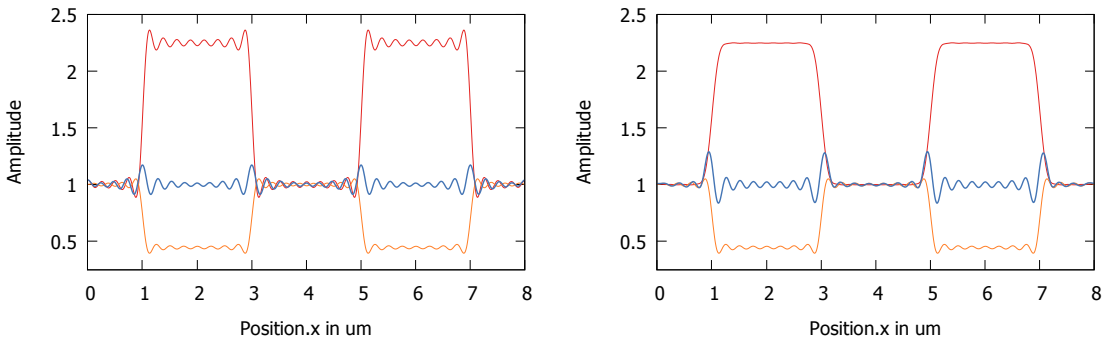
Since the Gibbs phenomenon only occurs for the field components, which are oriented normal to the grating grooves, the convergence speeds should not deviate a lot, when applying an apodization function to the permittivity function under TE polarized light. This can indeed be observed in Fig. 2.12b.



(a) TM polarization

(b) TE polarization

Figure 2.12: Comparison of the convergence speeds of the standard RCWA, the standard approach with inverse rule applied and with apodization.



(a) Without apodization

(b) With apodization

Figure 2.13: Reconstructed electric field (in orange), permittivity (in red) and electric displacement field (in blue) with (b) and without (a) apodization. In either case, the unphysical oscillations caused by the Laurent's rule remain.

Unfortunately though, apodization does also not improve the convergence speed for TM polarization either. Fig. 2.12b shows the convergence of the transmission efficiency of the 0th diffraction order for the same metallic grating that was analyzed before in the example of Sec. 2.5.3. The reason for the inferior convergence speed lies in the fact that the apodization changes the form of the design grating and with that also the target function of the Fourier expansion. Since the coefficients of the original Fourier expansion already corresponded to a reconstructed grating with the smallest square error based on the Fourier base functions[22],

a change in the grating form leads to new Fourier coefficients, which now correspond to a reconstructed grating with a greater square deviation from the original design. This is reflected in the convergence speed of the diffraction coefficients and can even be observed in the TE case (cf. Fig. 2.12b). Fig. 2.13b illustrates that the apodization also does not cure the critical situation at the concurrent jump discontinuities, but in fact increases the erroneous discontinuity in the displacement field.
It also does not seem promising to use a different apodization function, since the presented Gaussian apodization function in Eq. 2.46 already removes the over- and undershoots convincingly, while maintaining the steepness of the original grating slopes.

2.7 Local permeability

Most optical materials only have a very weak magnetic coupling component. Their permeability μ is assumed to be close to the one of the vacuum. Hence, their refractive index $n = \sqrt{\epsilon\mu}$ is only determined by the permittivity ϵ. However, in 1967 Veselago published his theoretical work about materials with negative permittivity and permeability[128], which laid the foundation of today's research about so called metamaterials. Metamaterials are artificial materials with subwavelength structures that act as simple electric components like nanowires or split ring resonators[78]. Excited by an electromagnetic field they can show custom electric and magnetic responses and that way emulate material properties that do not even exist in nature. Most popular are materials with negative values of the permeability and permittivity. They exhibit various counterintuitive properties like a negative index of refraction and a propagation vector of the electromagnetic field that points in opposite direction of the Poynting Vector, resulting in an antiparallel phase velocity. In 2000 Pendry proposed the use of a negative index material for the construction of a perfect lens[100]. By compensating for the exponential decay of evanescent waves with a negative index such a lens could reconstruct an image beyond the diffraction limit. The concept was first realized and confirmed by Melville and Blaikie[84] in 2005. Despite the promising perspective, major difficulties in a practical design and manufacturing process remain. Building the small structures that are needed for an artificial negative index material to work in the visible range is still a challenge. An even greater issue of all current realizations is the high dispersion and absorption that comes along with the negative index property. These facts will hopefully not prevent people from finding new materials and studying their nature. Thus, an RCWA algorithm that also covers non-trivial ($\mu \neq 1$), locally varying permeability is evidently advantageous and was also implemented in this work. It should, however, be noted that a similar suggestion was already made by Li[74] in 2003.

Since the underlying structure of metamaterials remain hidden to the incident light, they can be described in good approximation only by their macroscopic permittivity and permeability. Hence, in order to analyze dielectric and magnetic diffraction gratings also the permeability has to be modeled with a Fourier expansion (analogous to the permittivity in Eq. 2.1). The key difference in the algorithm again concerns the matrices $\mathbf{M}_1$ and $\mathbf{M}_2$ in the partial differential equation 2.14. And so, an equal treatment of the permittivity $\epsilon(x, y)$

and permeability $\mu(x, y)$ leads to a very symmetric form:

$$\mathbf{M}_1 := ik_0 \left[\begin{pmatrix} \mathbf{K_x}[\![\epsilon]\!]^{-1}\,\mathbf{K_y} & -\mathbf{K_x}[\![\epsilon]\!]^{-1}\mathbf{K_x} \\ \mathbf{K_y}[\![\epsilon]\!]^{-1}\,\mathbf{K_y} & -\mathbf{K_y}[\![\epsilon]\!]^{-1}\mathbf{K_x} \end{pmatrix} + \begin{pmatrix} 0 & \lceil\lfloor\mu\rfloor\rceil \\ -\lfloor\lceil\mu\rceil\rfloor & 0 \end{pmatrix} \right] \tag{2.47a}$$

$$\mathbf{M_2} := ik_0 \left[\begin{pmatrix} -\mathbf{K_x}[\![\mu]\!]^{-1}\mathbf{K_y} & \mathbf{K_x}[\![\mu]\!]^{-1}\mathbf{K_x} \\ -\mathbf{K_y}[\![\mu]\!]^{-1}\mathbf{K_y} & \mathbf{K_y}[\![\mu]\!]^{-1}\mathbf{K_x} \end{pmatrix} + \begin{pmatrix} 0 & -\lceil\lfloor\epsilon\rfloor\rceil \\ \lfloor\lceil\epsilon\rceil\rfloor & 0 \end{pmatrix} \right] \tag{2.47b}$$

2.8 Surface reliefs and multilayer gratings

Three-dimensional diffraction structures are not treated natively by the RCWA algorithm. Instead the z dependence of the grating's permittivity is always assumed to be constant and a Fourier expansion is only carried out in lateral direction (cf. Eq. 2.1). In order to analyze two-dimensional surface-relief gratings in the xz-plane, Kim et al.[63] published an approach called pseudo-Fourier model analysis (PFMA) that performs an additional Pseudo-Fourier expansion perpendicular to the grating. However, this approach leads to a considerable larger eigenvalue problem and thus significantly raise the computational effort.
The by far most common way to treat z structured gratings is the multilayer approximation, which was first suggested by Peng et al.[101] in 1975. The idea is to approximate the z dependence of the grating by a stack of N thin, planar-grating layers, each with averaged constant permittivity values along z.

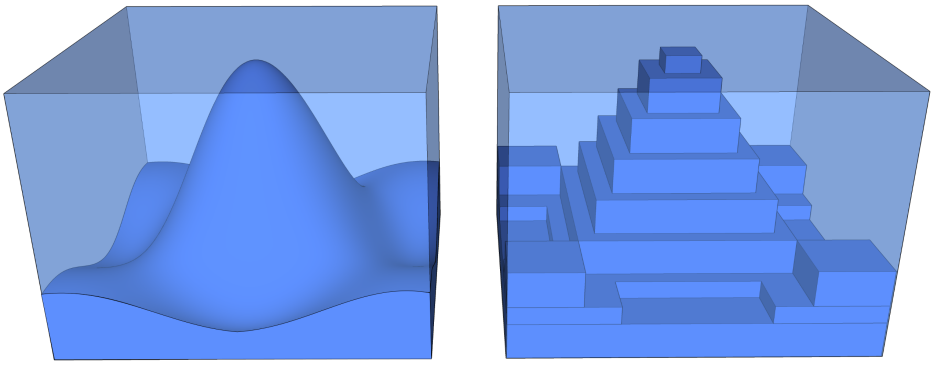


Figure 2.14: Multilayer approximation of a thick volume grating

The diffraction problem is then solved for each layer in terms of its eigenmodes and with the boundary conditions at each layer interface. This concept was first adopted to the RCWA by Moharam and Gaylord[87] in 1982. From the outer and all intermediate boundary conditions they derived a single big system of linear equations in order to calculated the diffraction coefficients $\mathbf{R}_\perp, \mathbf{T}_\perp$ and also the remaining unknowns of each layer l by Gaussian elimination. The T-matrix[7] approach is a more compact and efficient method. However, it

[7] T-Matrix Approach is an abbreviation for the term transfer matrix approach.

suffers from significant instability due to the inversion of very small values of the occurring exponentially damped evanescent waves. After first attempts to treat the problem by an artificial clipping of the exponential growth[120], further development led to intrinsic stable solutions like the R-matrix[8] approach[75], the S-matrix[9] approach[72, 73] and the enhanced transmission matrix approach (ETMA)[92] that avoid the problem of exponential growth by proper normalization.

2.8.1 T-Matrix Approach

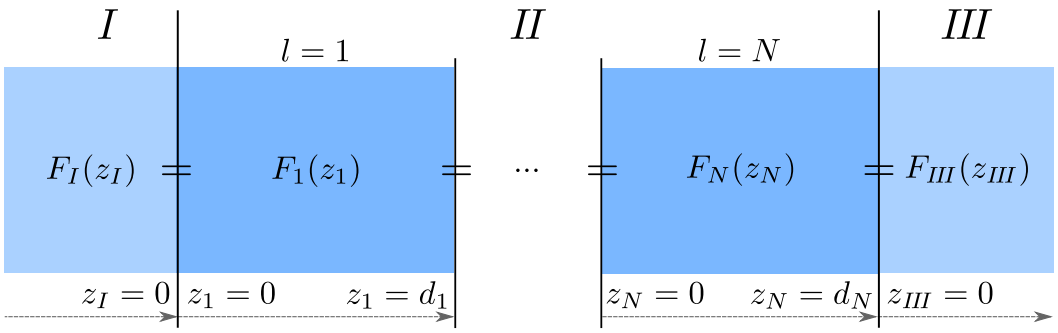


Figure 2.15: Electromagnetic fields with locally defined z-positions and matching boundary conditions at each layer interface

From Eq. 2.29 and $F_l(d) = F_{l+1}(0)$ (cf. Fig. 2.15) the recursion formula

$$\mathbf{F}_{l+1}(0) = \underbrace{\left[\mathbf{Q}_l \mathbf{P}_l(d) \mathbf{P}_l(0)^{-1} \mathbf{Q}_l^{-1}\right]}_{\text{Transfer Matrix } \mathbb{t}_l} \cdot \mathbf{F}_l(0) \tag{2.48}$$

can be derived, in which a transfer matrix $\mathbb{t}_l$ describes the propagation of the electromagnetic field across a single grating layer $l = 1..N$. The fields on both sides of a multilayer-grating can then be linked together by a transfer matrix $\mathbb{T}_N$ with the following definitions:

$$\mathbb{T}_l = \mathbb{T}_{l-1} \cdot \mathbb{t}_l,\ \mathbb{T}_0 = \mathbf{I} \tag{2.49a}$$

$$\Rightarrow \mathbb{T}_N = \prod_{l=N}^{1} \mathbb{t}_l \tag{2.49b}$$

$$\Rightarrow \mathbf{F}_{III}(0) = \mathbb{T}_N \cdot \mathbf{F}_I(0) \tag{2.49c}$$

Although in this form the T-matrix approach appears very intuitive and elegant, the method suffers severe numerical instabilities. They become apparent not until

[8] The R-matrix describes the ratio of the tangential components to the electric field and the tangential components of the magnetic field between different interfaces[73].

[9] S-matrix approach is an abbreviation for scattering matrix approach.

a full substitution of the variables in Eq. 2.50:

$$\underbrace{\left[\underbrace{\begin{pmatrix}\mathbf{I} & \mathbf{I}\\ \mathbf{C} & -\mathbf{C}\end{pmatrix}}_{\mathbf{Q}_{III}}\underbrace{\begin{pmatrix}\mathbf{T}_\perp\\ 0\end{pmatrix}}_{\mathbf{\Psi}_{III}}\right]_{III}}_{\mathbf{F}_{\mathbf{III}}(0)} = \underbrace{\left[\underbrace{\begin{pmatrix}\mathbf{V} & \mathbf{V}\\ \mathbf{W} & -\mathbf{W}\end{pmatrix}}_{\mathbf{Q}_l}\underbrace{\begin{pmatrix}\mathbf{X} & 0\\ 0 & \mathbf{X}^{-1}\end{pmatrix}}_{\mathbf{P}_l}\underbrace{\begin{pmatrix}\mathbf{V} & \mathbf{V}\\ \mathbf{W} & -\mathbf{W}\end{pmatrix}^{-1}}_{\mathbf{Q}_l^{-1}}\right]_l}_{\text{Transfer Matrix } \mathbb{t}_l}\underbrace{\left[\underbrace{\begin{pmatrix}\mathbf{I} & \mathbf{I}\\ \mathbf{C} & -\mathbf{C}\end{pmatrix}}_{\mathbf{Q}_I}\underbrace{\begin{pmatrix}\mathbf{L}_\perp\\ \mathbf{R}_\perp\end{pmatrix}}_{\mathbf{\Psi}_I}\right]_I}_{\mathbf{F}_{\mathbf{I}}(0)}. \tag{2.50}$$

Remark: Before analyzing the shortcomings of this approach, it should be noted that Eq. 2.50 reveals some insights to the RCWA algorithm and should therefore be reviewed briefly. Given is a detailed description of how the external fields on both sides of a single grating layer are connected through a transfer matrix $\mathbb{t}_l$ with $|\mathbb{t}_l| = 1$ that describes the wave coupling inside the grating (from right to left): The right site shows the tangential components of the Fourier coefficients of the electric field in the incident region. Matrix $\mathbf{Q}_I$ appends the magnetic counterparts. Describing a full electromagnetic field, the equation transcends the entrance interface to the structured layer obeying the continuity conditions. Here, the Fourier modes couple. By applying the matrix $\mathbf{Q}_l^{-1}$ the Fourier modes are transformed into uncoupled electric eigenmodes. Matrix $\mathbf{P}$ describes their undisturbed propagation through the grating to the exit interface, where they are finally transformed back into electric and magnetic Fourier modes by the matrix $\mathbf{Q}_l$. The resulting modes again match the Fourier modes on the outside.

The problem of this straight forward approach occurs in the process of inverting the diagonal matrix $\mathbf{X}$ (cf. $X_{q,q} = e^{-k_0\sqrt{\lambda_q}d}$), which describes the "propagation" of not only propagative, but also evanescent waves in the grating region. Since evanescent waves decay exponentially in distance to the boundaries, the entries in $\mathbf{X}$ may reach very small values and consequently can lead to arithmetic overflow, when inverting the matrix.

2.8.2 R- and S-matrix approach

The R- and the S-matrix approach are two similar but stable recursive algorithms. They solve Eq. 2.49c by using clever rearrangement schemes to avoid the potentially dangerous inversion of the matrix $\mathbf{X}_l$. The starting point is the T-matrix approach, which can be expressed either in terms of two orthogonally polarized eigenmodes $\mathbf{S}_\perp$ and $\mathbf{U}_\perp$

$$\mathbf{F}_{l+1}(0) = \underbrace{\left[\mathbf{Q}_l\mathbf{P}_l(d)\mathbf{P}_l(0)^{-1}\mathbf{Q}_l^{-1}\right]}_{\mathbb{t}_l}\cdot\mathbf{F}_l(0) \tag{2.51a}$$

$$\begin{pmatrix}\mathbf{S}_{l+1}\\ \mathbf{U}_{l+1}\end{pmatrix} = \mathbb{t}_l\begin{pmatrix}\mathbf{S}_l\\ \mathbf{U}_l\end{pmatrix},\quad \begin{pmatrix}\mathbf{S}_{l+1}\\ \mathbf{U}_{l+1}\end{pmatrix} = \mathbb{T}_l\begin{pmatrix}\mathbf{S}_0\\ \mathbf{U}_0\end{pmatrix} \tag{2.51b}$$

or by forwards and backwards propagating waves $\mathbf{t}_\perp$ and $\mathbf{r}_\perp$ (cf. 2.29)

$$\mathbf{\Psi}_{l+1} = \underbrace{\left[\mathbf{Q}_{l+1}^{-1}\mathbf{Q}_l\mathbf{P}_l(d)\mathbf{P}_l(0)^{-1}\right]}_{\mathbb{t}_l'}\mathbf{\Psi}_l \tag{2.52a}$$

$$\begin{pmatrix}\mathbf{t}_{l+1}\\ \mathbf{r}_{l+1}\end{pmatrix} = \mathbb{t}_l'\begin{pmatrix}\mathbf{t}_l\\ \mathbf{r}_l\end{pmatrix}, \quad \begin{pmatrix}\mathbf{t}_{l+1}\\ \mathbf{r}_{l+1}\end{pmatrix} = \mathbb{T}_l'\begin{pmatrix}\mathbf{t}_0\\ \mathbf{r}_0\end{pmatrix}. \tag{2.52b}$$

Rearranging the equations in 2.51b and 2.52b leads directly to the R- and the S-matrix approaches:

$$\text{R-Matrix Approach:} \quad \begin{pmatrix}\mathbf{S}_{l+1}\\ \mathbf{S}_l\end{pmatrix} = \mathbb{r}_l\begin{pmatrix}\mathbf{U}_{l+1}\\ \mathbf{U}_l\end{pmatrix}, \quad \begin{pmatrix}\mathbf{S}_{l+1}\\ \mathbf{S}_0\end{pmatrix} = \mathbb{R}_l\begin{pmatrix}\mathbf{U}_{l+1}\\ \mathbf{U}_0\end{pmatrix} \tag{2.53a}$$

$$\text{S-Matrix Approach:} \quad \begin{pmatrix}\mathbf{r}_{l+1}\\ \mathbf{t}_l\end{pmatrix} = \mathbb{s}_l\begin{pmatrix}\mathbf{r}_l\\ \mathbf{t}_{l+1}\end{pmatrix}, \quad \begin{pmatrix}\mathbf{r}_{l+1}\\ \mathbf{t}_0\end{pmatrix} = \mathbb{S}_l\begin{pmatrix}\mathbf{r}_0\\ \mathbf{t}_{l+1}\end{pmatrix} \tag{2.53b}$$

Their superior numerical stability is revealed in the corresponding recursion formulas

$$\mathbb{R}_l = \begin{pmatrix}\mathbb{r}_l^{11} - \mathbb{r}_l^{12}\left[\mathbb{r}_l^{22} - \mathbb{R}_{l-1}^{11}\right]^{-1}\mathbb{r}_l^{21} & \mathbb{r}_l^{12}\left[\mathbb{r}_l^{22} - \mathbb{R}_{l-1}^{11}\right]^{-1}\mathbb{R}_{l-1}^{12}\\ -\mathbb{R}_{l-1}^{21}\left[\mathbb{r}_l^{22} - \mathbb{R}_{l-1}^{11}\right]^{-1}\mathbb{r}_l^{21} & \mathbb{R}_{l-1}^{22} + \mathbb{R}_{l-1}^{21}\left[\mathbb{r}_l^{22} - \mathbb{R}_{l-1}^{11}\right]^{-1}\mathbb{R}_{l-1}^{12}\end{pmatrix} \text{ and} \tag{2.54a}$$

$$\mathbb{S}_l = \begin{pmatrix}\mathbb{s}_l^{11}\left[\mathbf{I} - \mathbb{S}_{l-1}^{12}\mathbb{s}_l^{21}\right]^{-1}\mathbb{S}_{l-1}^{11} & \mathbb{s}_l^{12} + \mathbb{s}_l^{11}\mathbb{S}_{l-1}^{12}\left[\mathbf{I} - \mathbb{s}_l^{21}\mathbb{S}_{l-1}^{12}\right]^{-1}\mathbb{s}_l^{22}\\ \mathbb{S}_{l-1}^{21} + \mathbb{S}_{l-1}^{22}\mathbb{s}_l^{21}\left[\mathbf{I} - \mathbb{S}_{l-1}^{12}\mathbb{s}_l^{21}\right]^{-1}\mathbb{S}_{l-1}^{11} & \mathbb{S}_{l-1}^{22}\left[\mathbf{I} - \mathbb{s}_l^{21}\mathbb{S}_{l-1}^{12}\right]^{-1}\mathbb{s}_l^{22}\end{pmatrix}, \tag{2.54b}$$

in which each two-digit upper number indicates a block matrix element. All matrices in Eq. 2.54a and Eq. 2.54b that might be ill conditioned and are going to be inverted are subtracted off a less critical matrix and are thereby normalized. For example, if the product $S_{l-1}^{12}s_l^{21}$ in $\mathbb{S}_l^{11}$ contained values of very small magnitude, the inversion of the matrix $\left[\mathbf{I} - S_{l-1}^{12}s_l^{21}\right]$ would be harmless. However, since the R-matrix is not normalized with a unit matrix, cases might occur, where the R-matrix approach is still unstable. The complete derivation and comparison of the R- and S-matrix approach was published by Li[73].

2.8.3 Enhanced Transmission Matrix Approach

The enhanced transmission matrix approach (ETMA) is another efficient and numerically stable approach to simulate multi layer gratings. It was first introduced by Moharam and Gaylord as a partial solution[92] and later completed by Tan[123]. Due to its superior stability and relatively simple usage, it is also used in all following simulations of this work.

The continuity condition in Eq. 2.30 and 2.31 can be generalized to connect every two adjacent fields:

$$\begin{pmatrix} \mathbf{M}_l^{11} & \mathbf{M}_l^{12} \\ \mathbf{M}_l^{21} & -\mathbf{M}_l^{22} \end{pmatrix} \begin{pmatrix} \mathbf{t}_l \\ \mathbf{r}_l \end{pmatrix} = \begin{pmatrix} \mathbf{M}_{l+1}^{12} & \mathbf{M}_{l+1}^{11} \\ \mathbf{M}_{l+1}^{22} & -\mathbf{M}_{l+1}^{21} \end{pmatrix} \begin{pmatrix} \mathbf{t}_{l+1} \\ \mathbf{r}_{l+1} \end{pmatrix} \tag{2.55}$$

Therein, the matrix elements are defined depending on the specific region they refer to. Thus, in region *II* ($1 \leq l \leq N$) it is

$$\mathbf{M}_l^{11} := \mathbf{VX} \tag{2.56a}$$
$$\mathbf{M}_l^{12} := \mathbf{V} \tag{2.56b}$$
$$\mathbf{M}_l^{21} := \mathbf{WX} \tag{2.56c}$$
$$\mathbf{M}_l^{22} := \mathbf{W} \tag{2.56d}$$

and in region *I* ($l = 0$) and *III* ($l = N + 1$) it is

$$\mathbf{M}_l^{11} := \mathbf{I} \tag{2.57a}$$
$$\mathbf{M}_l^{12} := \mathbf{I} \tag{2.57b}$$
$$\mathbf{M}_l^{21} := \mathbf{C}_l \tag{2.57c}$$
$$\mathbf{M}_l^{22} := \mathbf{C}_l. \tag{2.57d}$$

With the ansatz $\mathbf{a}_l \mathbf{t}_l = \mathbf{r}_l$ Eq. 2.55 can be rearranged to

$$\begin{pmatrix} \mathbf{r}_l \\ \mathbf{t}_{l+1} \end{pmatrix} = \begin{pmatrix} \mathbf{M}_l^{12} & \mathbf{f}_{l+1} \\ -\mathbf{M}_l^{22} & \mathbf{g}_{l+1} \end{pmatrix}^{-1} \begin{pmatrix} -\mathbf{M}_l^{11} \\ \mathbf{M}_l^{21} \end{pmatrix} \mathbf{t}_l =: \begin{pmatrix} \mathbf{a}_l \\ \mathbf{b}_l \end{pmatrix} \mathbf{t}_l, \tag{2.58}$$

in which $\mathbf{a}_l$, $\mathbf{b}_l$ and $\mathbf{f}_l$, $\mathbf{g}_l$ are related by the recursion formula

$$\begin{pmatrix} \mathbf{a}_l \\ \mathbf{b}_l \end{pmatrix} = \begin{pmatrix} \mathbf{M}_l^{12} & \mathbf{f}_{l+1} \\ -\mathbf{M}_l^{22} & \mathbf{g}_{l+1} \end{pmatrix}^{-1} \begin{pmatrix} -\mathbf{M}_l^{11} \\ -\mathbf{M}_l^{21} \end{pmatrix} \tag{2.59a}$$
$$\begin{pmatrix} \mathbf{f}_l \\ \mathbf{g}_l \end{pmatrix} = \begin{pmatrix} -\mathbf{M}_l^{12} - \mathbf{M}_l^{11}\mathbf{a}_l \\ \mathbf{M}_l^{22} - \mathbf{M}_l^{21}\mathbf{a}_l \end{pmatrix}. \tag{2.59b}$$

For region *II* ($1 \leq l \leq N$), this results in the explicit form

$$\begin{pmatrix} \mathbf{a}_l \\ \mathbf{b}_l \end{pmatrix} = \begin{pmatrix} \mathbf{V}_l & \mathbf{f}_{l+1} \\ -\mathbf{W}_l & \mathbf{g}_{l+1} \end{pmatrix}^{-1} \begin{pmatrix} -\mathbf{V}\mathbf{X}_l \\ -\mathbf{W}\mathbf{X}_l \end{pmatrix} \tag{2.60a}$$

$$\begin{pmatrix} \mathbf{f}_l \\ \mathbf{g}_l \end{pmatrix} = \begin{pmatrix} -\mathbf{V}_l \left[\mathbf{I} + \mathbf{X}\mathbf{a}\right] \\ \mathbf{W}_l \left[\mathbf{I} - \mathbf{X}\mathbf{a}\right] \end{pmatrix}. \tag{2.60b}$$

The recursion formula starts with defined quantities in the transmitting region ($l = N + 1$):

$$\begin{pmatrix} \mathbf{f}_{N+1} \\ \mathbf{g}_{N+1} \end{pmatrix} = \begin{pmatrix} -\mathbf{I} \\ \mathbf{C}_{III} \end{pmatrix} \tag{2.61}$$

and ends with the results for the reflection coefficients in the incident region ($l = 0$) through

$$\begin{pmatrix} \mathbf{R} \\ \mathbf{t}_1 \end{pmatrix} = - \begin{pmatrix} \mathbf{I} & \mathbf{f}_1 \\ -\mathbf{C}_l & \mathbf{g}_1 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{I} \\ \mathbf{C}_l \end{pmatrix} \mathbf{L}. \tag{2.62}$$

The transmission coefficients can then be obtained by following the recursion formula of $\mathbf{b}$ derived from Eq. 2.58:

$$\mathbf{t}_{l+1} = \mathbf{b}_l \mathbf{t}_l \tag{2.63a}$$

$$\mathbf{t}_{N+1} = \left(\mathbf{b}_N \mathbf{b}_{N-1} ... \mathbf{b}_2 \mathbf{b}_1\right) \mathbf{t}_0 \tag{2.63b}$$

$$\mathbf{T} = \left(\prod_{l=N}^{1} \mathbf{b}_l \right) \mathbf{L} \tag{2.63c}$$

The reason for the superior numerical stability of the ETMA over the T-matrix approach lies in the construction of the block matrix $\begin{pmatrix} \mathbf{f}_l & \mathbf{g}_l \end{pmatrix}^T$ in Eq. 2.60b. In it the diagonal propagation matrix $\mathbf{X}$ (with $X_{q,q} = e^{-k_0 \sqrt{\lambda_q} d}$), which can reach very small values for damped exponential fields, is not inverted directly. Instead it is first normalized by adding a unit matrix and then inverted in Eq. 2.60b. Thus any numerically unstable inversion is avoided even for thick layers.

2.9 Diffraction order efficiencies

The diffraction order efficiency is an important quantity when studying diffraction gratings. It describes the amount of power that is carried in each diffraction order averaged over time. Gratings are often optimized for high or low efficiency values in certain orders. Blaze gratings for example, which are used in many monochromators, are designed to channel most of the incident energy in a single diffraction order, while other orders are suppressed.
Calculating the order-specific efficiencies of a grating requires an energy analysis. If the surface of a diffraction grating is illuminated, the amount of energy dW that flows through a small area element $d\mathbf{A}$ in one time unit dt can be determined by Poynting theorem:

$$dW = \mathbf{S} \cdot dt \cdot d\mathbf{A} \tag{2.64a}$$

$$\Rightarrow \langle dW \rangle = \langle \mathbf{S} \rangle \cdot d\mathbf{A}. \tag{2.64b}$$

Since the RCWA only considers steady state systems, the quantity of interest is the time averaged energy. It is defined by the time averaged Poynting vector $\langle \mathbf{S} \rangle$ (cf. A.3), which describes the directional energy flux, times the surface vector element $d\mathbf{A}$ of the irradiated area. Since the total energy must always be positive, both vectors shall have the same orientation. The surface vector element can be expressed by the product of the area dA and the surface normal of this area element:

$$d\mathbf{A} = \mathbf{n} dA \tag{2.65}$$

Due to the law of energy conservation, all power that flows into a non-absorbing optical element must either be reflected or transmitted (cf. Fig. 2.16). This can be expressed by

$$\langle \mathbf{S}_i \rangle d\mathbf{A}_i = \sum_{m,n} \left[\langle \mathbf{S}_r \rangle d\mathbf{A}_r \right]_{m,n} + \sum_{m,n} \left[\langle \mathbf{S}_t \rangle d\mathbf{A}_t \right]_{m,n} \tag{2.66a}$$

$$\Rightarrow 1 = \frac{\sum \left[\langle \mathbf{S}_r \rangle d\mathbf{A}_r \right]_{m,n}}{\langle \mathbf{S}_i \rangle d\mathbf{A}_i} + \frac{\sum \left[\langle \mathbf{S}_t \rangle d\mathbf{A}_t \right]_{m,n}}{\langle \mathbf{S}_i \rangle d\mathbf{A}_i}, \tag{2.66b}$$

where $\left[\langle \mathbf{S}_\rho \rangle d\mathbf{A}_\rho \right]_{m,n}$ with $\rho = r, t$ describes the power flow of a single mode in the far field with index (m, n). The reflected and transmitted power per mode in relation to the power input can then be defined by the diffraction order efficiencies $\eta^R_{m,n}$ and $\eta^T_{m,n}$ respectively:

$$\eta^R_{m,n} := \frac{\left[\langle \mathbf{S}_r \rangle d\mathbf{A}_r \right]_{m,n}}{\langle \mathbf{S}_i \rangle d\mathbf{A}_i} \tag{2.67a}$$

$$\eta^T_{m,n} := \frac{\left[\langle \mathbf{S}_t \rangle d\mathbf{A}_t \right]_{m,n}}{\langle \mathbf{S}_i \rangle d\mathbf{A}_i} \tag{2.67b}$$

The time averaged Poynting vector can be expressed[10] by

$$\langle \mathbf{S} \rangle = \frac{1}{2Z_0 k_0 \mu_l} |\mathbf{E}|^2 Re\left\{\mathbf{k}\right\}, \tag{2.68}$$

where $\mathbf{E}$ describes the electric field and $\mathbf{k}$ is the wavevector, which, for a positive refractive index, points in the same direction as the Poynting vector. Identifying the (m,n)-th discrete diffraction orders, the transmission efficiency, as an example, can then be written as

$$\eta^T_{m,n} = \frac{\frac{1}{2Z_0 k_0 \mu_{III}} |\mathbf{E}_{t;m,n}|^2 Re\left\{\mathbf{k}_{t;m,n}\right\} \mathbf{n}_t}{\frac{1}{2Z_0 k_0 \mu_I} |\mathbf{E}_i|^2 Re\left\{\mathbf{k}_i\right\} \mathbf{n}_i} = |T_{m,n}|^2 Re\left\{\frac{\mathbf{k}_{t;m,n}\mathbf{n}_t}{\mathbf{k}_i \mathbf{n}_i}\right\} \frac{\mu_I}{\mu_{III}}. \tag{2.69}$$

The permeability μ_l of the outer regions $l = I, III$ is considered to be a constant, real value. And since the surface normals were defined to point in the same direction along z as the power flow and the wave vector, it is $\mathbf{k}_i \mathbf{n}_i = k_{i;z}$, $\mathbf{k}_r \mathbf{n}_r = k_{r;z} = \gamma_I$ and $\mathbf{k}_t \mathbf{n}_t = k_{t;z} = \gamma_{III}$, and the area elements dA cancel out.

The diffraction efficiencies can now be specified for the individual transmission and reflection orders as

$$\eta^T_{m,n} = |T_{m,n}|^2 Re\left\{\frac{\gamma_{III;m,n}}{k_{i;z}}\right\} \frac{\mu_I}{\mu_{III}} \tag{2.70a}$$

$$\eta^R_{m,n} = |R_{m,n}|^2 Re\left\{\frac{\gamma_{I;m,n}}{k_{i;z}}\right\}. \tag{2.70b}$$

And finally, the global amount of transmitted and reflected power is the sum over all diffraction orders:

$$\text{Global Transmission}: \quad \mathcal{T} := \sum_{m,n} \eta^T_{m,n} \tag{2.71a}$$

$$\text{Global Reflection}: \quad \mathcal{R} := \sum_{m,n} \eta^R_{m,n}, \tag{2.71b}$$

According to Moharam and Gaylord[88] energy conservation is always ensured for non-absorbing gratings and can be verified by the following relation:

$$1 = \mathcal{T} + \mathcal{R}. \tag{2.72}$$

[10] In the interest of clarity the exact derivation of the averaged Poynting vector is avoided here, but can be reviewed in Sec. 5.2.2, Eq. 5.19f.

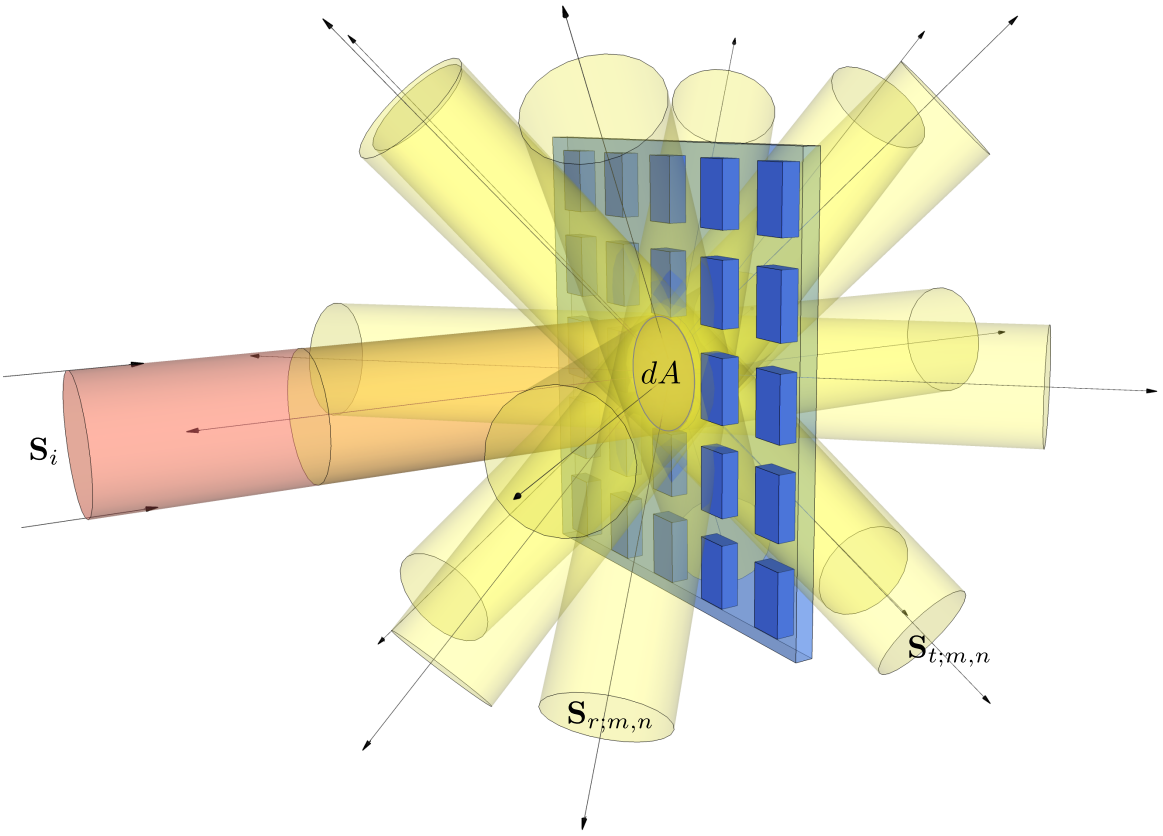


Figure 2.16: Schematic representation of an optical grating, which shows the power flow of the incident (red) and the diffracted light (yellow) in multiple diffraction orders.

Chapter 3

Localized input field approach

3.1 Motivation

Typically, when studying diffraction effects with rigorous methods, the incident light is assumed to be an infinite plane wave. This is true in many standard theories like Mie scattering, multilayer theory and also in the standard RCWA. However, in a practical application the light source might as well have a Gaussian profile. Other optical setups require focused beams and even plane waves are often distorted due to imperfections of the optical components. So there is an apparent need for an additional extension of the rigorous coupled-wave analysis (RCWA) to localized incident fields.

3.1.1 Previous works

Most previous theories are not able to describe the rigorous diffraction grating problem for a three-dimensional converging-diverging Gaussian beam at oblique, non-classical[1] incidence. As pointed out in Sec. 1.2.3, only few analytical solutions are known for special diffraction problems and none for arbitrary three-dimensional dielectric gratings. Finite incident fields like Gaussian beams or even arbitrarily shaped beams would only further increase the mathematical difficulty. Numerical approaches, on the other hand, may aim for a decomposition of the incident field into plane waves and superpose the results. However, this drastically increases the computational effort, since multiple rigorous calculations would have to be performed separately for every plane wave. Nevertheless, there have been multiple attempts to approach this problem in the rigorous regime.
1976 Chu and Tamir[23] used the plane wave decomposition to apply a cylindric Gaussian beam to Bragg diffraction theory. This way, they were the first to analyze the progression of a Gaussian beam inside a periodically modulated medium as well as the distortion effects on the profiles of the transmitted and diffracted beam. In 1980 Moharam and Gaylord[90] also applied a two-dimensional (cylindric) Gaussian beam to the theory of Bragg diffraction using Solymar's[116] two-dimensional extension of Kogelnik's famous coupled-wave method[66]. The analyses therefore also only considered the transmitted beam and a single diffracted beam as the only two diffraction orders and neglected the second partial derivatives in the wave equation. The study investigates diffraction efficiencies as a function of different system parameters. A very detailed overview of further studies is given by Wu and Gaylord[133],

[1] For one-dimensional gratings this is referred to as conical incidence (cf. Sec. 1.2.1).

including approaches for homogeneous multilayer structures and the scalar regime.
However, all these methods are restricted to the Bragg-diffraction regime, where higher diffraction orders are neglected. And it was not before 2005, when Wu and Gaylord[133] presented a method called finite-beam RCWA (FB-RCWA), which was able to rigorously analyze three-dimensional converging–diverging Gaussian beam diffraction by a volume grating. In order to model the Gaussian beam, they applied the plane wave decomposition (in this work referred to as Angular Spectrum Decomposition (ASD)), successively calculated the full RCWA for every incident plane wave and superposed the results. Due to the nature of the ASD, this method is not restricted to Gaussian beams, but can generally be applied to any arbitrary wave front. But as mentioned earlier, the downside of this rather straight forward approach is that the ASD acts like a multiplicator of the computational effort, since every decomposed plane wave requires the full calculation of the standard RCWA. This significantly limits the practical use of the Finite Beam method.

3.1.2 Novel approach: the LIF-RCWA

In this work, a new method is presented that can handle finite beams and arbitrary wavefronts for the three-dimensional diffraction grating problem. This method is called localized input field RCWA (LIF-RCWA) and enables the rigorous simulation of focusing problems, while supporting extended flexibility in polarization as it is not limited to pure linear and circular polarization anymore. The method uses the angular spectrum decomposition (ASD) (cf. Sec. 1.1.6) to model the incident field and an adjusted version of the standard RCWA algorithm to calculate the diffraction coefficients. In contrast to previous methods the computing effort of the LIF-RCWA remains equal to the standard RCWA algorithm for one plane wave. A substantial part of this concept was recently published in [6].

3.2 The idea

In the standard form of the rigorous coupled-wave analysis (RCWA), the propagating light inside a diffractive structure as well as the reflected and the transmitted light is already modeled as a superposition of plane waves. The incident light, however, is always only a single plane wave. The main idea of the LIF-RCWA is to dissolve this asymmetry and to natively excite multiple plane waves in the incident light.
Eq. 3.1 shows the standard representation of the incident light as a sparse vector $\mathbf{L}$ (cf. Eq. 2.26a)

$$\mathbf{L}_x = \begin{pmatrix} 0 \\ \vdots \\ E_{1;x} \\ \vdots \\ 0 \end{pmatrix}, \mathbf{L}_y = \begin{pmatrix} 0 \\ \vdots \\ E_{1;y} \\ \vdots \\ 0 \end{pmatrix}, \mathbf{L}_z = \begin{pmatrix} 0 \\ \vdots \\ E_{1;z} \\ \vdots \\ 0 \end{pmatrix}. \tag{3.1}$$

It is only populated with the zeroth mode for excitation, where $\mathbf{E_1}$ is the normalized electric field amplitude.

The goal of the new approach is to modify the standard RCWA in a way that it can propagate multiple input modes in a single calculation. This way, the incident vector could be filled with additional higher modes of the ASD of an arbitrary input light field in order to natively simulate a broad variety of light sources.

3.3 Preparation and preliminary considerations

In accordance to the definition of the reflected and the transmitted field in Eqs. 2.6, now also the electric incident field shall be expressed as a superposition of plane waves with a general vector $\mathbf{L} = (\mathbf{L}_x\ \mathbf{L}_y\ \mathbf{L}_z)^T$:

$$E_j(x,y,z) = \sum_{m,n} L_{j;m,n} e^{i\left(k_{x;m}x + k_{y;n}y + \gamma_{I;m,n}z\right)}, \qquad j = x,y,z \tag{3.2}$$

This definition offers great flexibility in modeling various input fields. And their rigorous propagation as a superposition of plane waves should be fundamentally valid due to the linear nature of the RCWA. However, as demonstrated in the following, the definition is still ambiguous and thereby raises the requirement for consistency.

An incident plane wave, for example, which is tilted in the x/z plane, can either be described by the selection of a certain non-zero mode m_q by

$$L_{m,n} = \delta_{m,m_q}\delta_{n,n_0}, \tag{3.3a}$$
$$k_{i;x} = k_0 n_I sin(\theta)cos(\phi) = 0, \ \Rightarrow\ k_{x;m} = 2\pi\, m/P_x \tag{3.3b}$$
$$k_{i;y} = k_0 n_I sin(\theta)sin(\phi) = 0, \ \Rightarrow\ k_{y;n} = 2\pi\, n/P_y \tag{3.3c}$$

or with an additional angular displacement $k_{i;x}$ to the zeroth mode:

$$L_{m,n} = \delta_{m,m_0}\delta_{n,n_0}, \tag{3.4a}$$
$$k_{i;x} = 2\pi m_q/P_x, \ \Rightarrow\ k_{x;m} = 2\pi\,(m+m_q)/P_x \tag{3.4b}$$
$$k_{i;y} = 0, \ \Rightarrow\ k_{y;n} = 2\pi\, n/P_y \tag{3.4c}$$

Eqs. 3.3a-3.3c describe a shift in k-space, whereas Eqs. 3.4a-3.4c describe a shift of the k-space coordinate system itself. Both definitions represent the same physical situation. And they both would in fact result in the same grating response in an untruncated calculation. However, in a numerical simulation with finite truncation limits, usually chosen symmetrically around the zeroth order, k-space is finite and the equality between the two approaches is not given naturally.

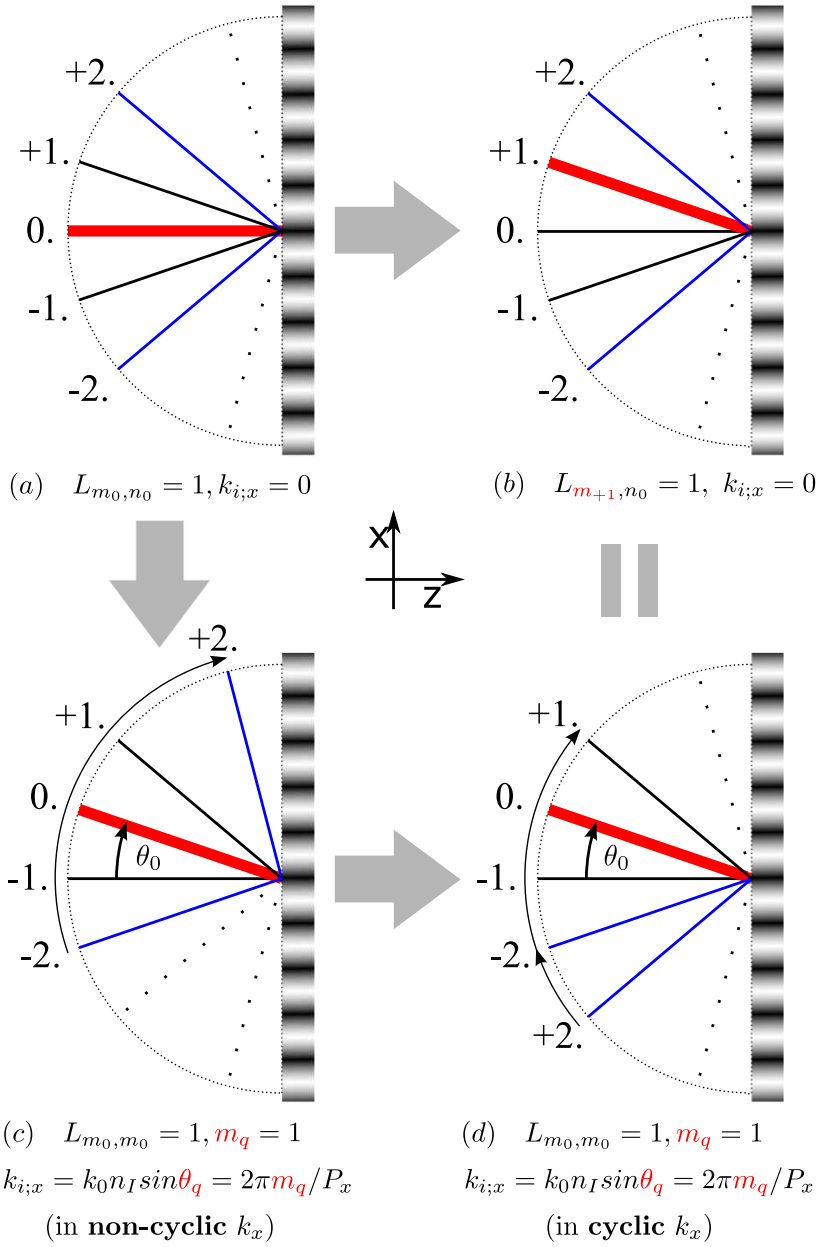


Figure 3.1: Two different ways to model a single tilted plane wave as an input field for the LIF-RCWA.

Fig. 3.1 depicts the different situations. The straight lines that point from the border of a half circle to the center of the grating represent the discrete k-space (or pupil) and all possible directional vectors of incident plane waves. The bold red lines indicate the wave vector of the only excited plane wave. The two blue lines at ±2th order indicate the highest untruncated modes in this example.

Fig. 3.1a shows the excitation of the zeroth mode (m_0, n_0) at normal incidence. In Fig. 3.1b a tilted incidence is achieved by exciting the plus first order mode (m_1, n_0) and Fig. 3.1c shows the same tilt achieved by a global shift of k-space itself. In fact, the two latter situations are the same for the incident light. However, if the propagation through the grating is performed in shifted k-space the resulting diffraction coefficients will differ, because of the different harmonics being involved.
The solution is a *cyclic shift* that wraps the modes from one end of the spectrum to the other. This is shown in Fig. 3.1c, where the remaining difference to Fig. 3.1b is only a matter of indexing.

3.4 Mode truncation of the EM-fields

In order to allow a global angular shift to all frequencies while obeying the cyclic wrap depicted in Fig. 3.1d, the tangential components of the wave vector need to be slightly redefined as

Cyclic definition:

$$k_{x;m} := \text{wrap}(k_{x;m}, -\pi\, M/P_x, +\pi\, M/P_x) \quad \text{and} \tag{3.5a}$$
$$k_{y;n} := \text{wrap}(k_{y;n}\,, -\pi\, N/P_y\,, +\pi\, N/P_y\,) \tag{3.5b}$$

with the length of vector $\mathbf{k}$ equal to $N \times M$ and

$$\text{wrap}(v, a, b) := v - (a - b)\left\lfloor \frac{v - a}{b - a} \right\rfloor. \tag{3.6}$$

The lower square bracket notation indicates the floor function. Its value is the largest integer number not greater than the input value inside the brackets. As previously shown in Eq. 2.8, k_z, the normal z component of the wave vector, is derived from the dispersion relation.

Discussion
The cyclic wrapping of the k-vectors introduces a periodicity in k-space, which corresponds to a discrete (or discretized) input field in position space. The consequences of this shall be discussed briefly, before continuing with further details of the method.
First of all, a discretized position space implies that the electromagnetic field is only known at discrete points. Thus the RCWA will only solve the diffraction problem on a discrete space grid. However, the Fourier coefficients of the grating permittivity are usually also obtained by a discrete Fourier transform (DFT), which assumes a periodic and discrete nature of the signal in both space and frequency domain. This is in fact a common procedure in digital signal processing. The necessary condition for an exact discrete representation of the original (continuous) signal is given by the Nyquist–Shannon sampling theorem. It states that if a bandlimited periodic signal is sampled at a rate more than twice its highest frequency over one

fundamental period, it can be exactly reconstructed from the samples. This also provides the proper cut off frequency for a mode truncation in the RCWA. For non-bandwidth limited signals though, the inevitable truncation of the infinite Fourier series otherwise act as a bandpass filtering operation that is applied to any grating structure and electromagnetic field analyzed by the RCWA.
The periodic frequency domain has yet another implications: The new definition of the wave vectors in Eq. 3.5a ensures formal consistency in the definition of a tilted incident wave. However, it should be avoided to actually occupy frequencies in the incident field above the given truncation limit, since this would violate Nyquist–Shannon sampling theorem and would result in aliasing effects: As a consequence of the periodic frequency domain, a single excited frequency above the cut-off frequency would appear on the opposite side of the spectrum as an incident plane wave, which is mirrored at the perpendicular of the grating interface at (z = 0). And this is most likely not intended. Also a localized input field, which consists of a multiple of frequencies will split into two parts of different directions, if the overall k-shift moves the field beyond the truncation limit. Thus, it should always be ensured, that the localized input field has still some extra distance to the truncation limits in k-space. This aspect is especially crucial since the convolution of the structure with the input field results in a broadening of the output spectrum.

3.5 Mode truncation of the grating permittivity

The next question concerns the form of the Toeplitz matrix in the RCWA, which appears in the convolution of the electric field and the permittivity in the frequency domain. In order to simplify matters, these considerations will be conducted for one-dimensional gratings. However, they are equally applicable to two-dimensional structures.

Starting point is the constitutive equation for the electric displacement field(cf. Eq. 1.2a):

$$D(x) = \epsilon(x)E(x) \tag{3.7a}$$

$$= \sum_{m=-\infty}^{+\infty} (\epsilon * \mathbf{S})_m \, e^{ik_{x;m}x} \tag{3.7b}$$

$$= \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \epsilon_{m-n} S_n e^{ik_{x;m}x} \tag{3.7c}$$

The product of permittivity $\epsilon(x)$ and the electric field $E(x)$ translates into a convolution of the Fourier coefficients ϵ_m and S_n.

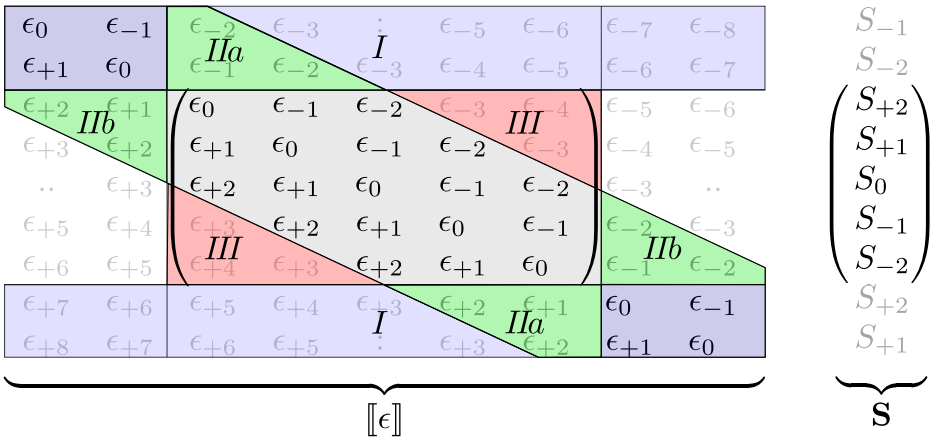


Figure 3.2: Exemplary matrix-vector notation of a discrete convolution between the permittivity ϵ and the electric field S in Fourier space. Against proper bandwidth limiting, the standard RCWA partially considers additional higher modes in the segments *III*, while ignoring necessary frequencies in the segments *IIa*.

As described in Eq. 2.12 and Fig. 2.2 such discrete convolution can also be written as a matrix-vector product, in which the matrix has Toeplitz form:

$$\epsilon * \mathbf{S} = [\![\epsilon]\!]\mathbf{S} \tag{3.8}$$

In an exact calculation the number of discrete frequencies might not be limited and thereby Eq. 3.8 would assume quantities of infinite size. However, the proper execution of the necessary truncation is not obvious.

The right-hand side of Eq. 3.8 is visualized in Fig. 3.2, where the convolution matrix $[\![\epsilon]\!]$ is divided into different segments of class $I - III$.
It is important to notice, that even a symmetric truncation of the infinite sums in Eq. 3.7c is not equal to a band limitation of $\epsilon(x)$. Such truncation reduces the infinite convolution matrix $[\![\epsilon]\!]$ to a square Toeplitz matrix of size $M \times M$ and the vector $\mathbf{S}$ to size $M = 2M_0 + 1$. This approach is used by the standard RCWA algorithm, but leads to some coefficients in $[\![\epsilon]\!]$ (cf. segments of class *III*), which are of higher order than the truncation limit, while neglecting lower order modes in the segments of class *IIa*. A proper band limitation of $\epsilon(x)$ on the other hand results in a different form, which is an infinite parallelogram around the main diagonal as indicated in gray in Fig. 3.2. All other entries in the matrix (cf. segments of class *III*) would be considered zero as they refer to higher modes, which are supposed to be neglected. Since the electric field vector $\mathbf{S}$ is assumed to be periodic in the LIF-approach (cf. Sec. 3.4), the matrix vector product $[\![\epsilon]\!]\mathbf{S}$ on the side of Eq. 3.8 can be written as an infinitely repeating vector of only M different terms. Thus, the truncation of the first sum over m in Eq. 3.7c now only discards redundancy. This is equal to setting all entries in the segments of class I to zero and leads to a non-square matrix of size $(2M_0 + 1) \times (4M_0 + 1)$. However, this form does not fit to the $(2M_0 + 1)$ entries of a truncated vector $\mathbf{S}$. If, on the other hand,

not only **S** is assumed periodic, but also the spectrum of $\epsilon(x)$, then writing out the full matrix-vector multiplication reveals that the protruding elements in the segments of class *IIb* re-emerge in the segments of class *III* - matching just the same coefficients of a periodic vector **S**. Thus, by mapping the entries of the segments of class *IIb* to the segments of class *III* the complete set of coefficients of the bandlimited permittivity function is condensed into a square matrix of size $(2M_0+1)\times(2M_0+1)$. This new matrix is called a circulant Toeplitz matrix and can be defined for either one- or two-dimensional gratings:

Cyclic definition for 1D-gratings:

$$[\overset{\infty}{\epsilon}]_{m,n} = \epsilon_{\mathrm{wrap}(m-n,0,M)} \tag{3.9a}$$

Example:

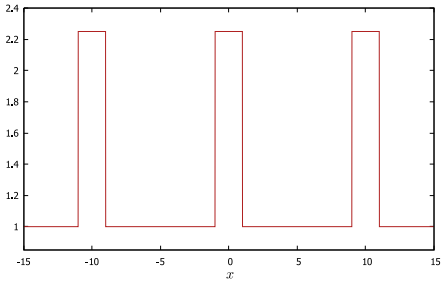
$$[\overset{\infty}{\epsilon}] = \begin{pmatrix} \epsilon_0 & \epsilon_{-1} & \epsilon_{-2} & \epsilon_{+2} & \epsilon_{+1} \\ \epsilon_{+1} & \epsilon_0 & \epsilon_{-1} & \epsilon_{-2} & \epsilon_{+2} \\ \epsilon_{+2} & \epsilon_{+1} & \epsilon_0 & \epsilon_{-1} & \epsilon_{-2} \\ \epsilon_{-2} & \epsilon_{+2} & \epsilon_{+1} & \epsilon_0 & \epsilon_{-1} \\ \epsilon_{-1} & \epsilon_{-2} & \epsilon_{+2} & \epsilon_{+1} & \epsilon_0 \end{pmatrix}$$

Cyclic definition for 2D-gratings:

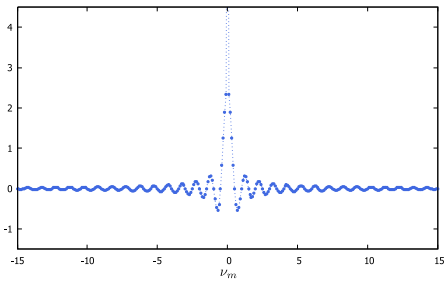
$$[\![\overset{\infty}{\epsilon}]\!]_{m\cdot M+q,n\cdot M+r} = \epsilon_{\substack{\mathrm{wrap}(m-n,\,0,\,M),\\ \mathrm{wrap}(\,q-r,0,N\,)}} \tag{3.9b}$$

Discussion

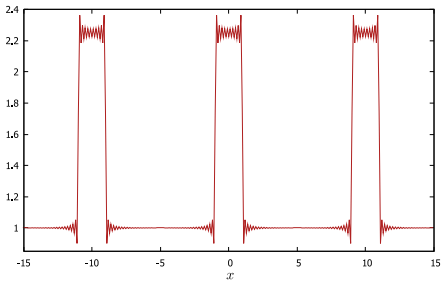
Without any bandlimiting, the permittivity and the electric field is described by continuous, periodic functions, which have an unbound, non-periodic spectrum as illustrated in Figs. 3.3a and 3.3b for a rect-like grating structure. Due to the infinite length of vector **S**, also the convolution matrix $[\epsilon]$ would be of infinite size (cf. Fig. 3.2). The latter would even be true when only the permittivity function was bandlimited as shown in Fig. 3.3d. The corresponding reconstructed permittivity for the exemplary grating is depicted in Fig. 3.3c. In the standard RCWA there are two common ways to obtain the Fourier coefficients of the permittivity. They can either be calculated by solving the Fourier integral (cf. Eq. 2.2) or by numerically calculating the discrete Fourier transform (cf. Eq. 2.3) with a preferably high number of sampling points - resulting in a large or even unbound spectrum. In both cases the convolution matrix $[\epsilon]$ is usually constructed by only a truncated subset of the spectrum in order to obtain a square matrix with manageable size. However, as illustrated in Fig. 3.3f, this truncation scheme leads to a convolution matrix in which all coefficients except the zero frequency are either underrepresented, as in the regions of the spectrum below the cutoff frequency (cf. segments of class *IIa*), or falsely accounted for as in the regions above the cutoff frequency (cf. segments of class *III*). Due to this inconsistency, it is not possible to derive any reasonable grating structure from the convolution matrix that is used in the standard approach. Thus its physical meaning can be called into question even though in the limit of a high truncation order the RCWA results are known to converge to the correct solutions.



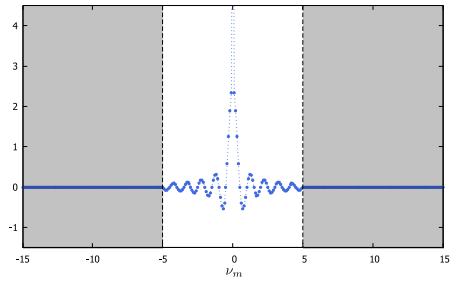
(a) Original, non-bandlimited $\epsilon(x)$: piecewise continuous and periodic



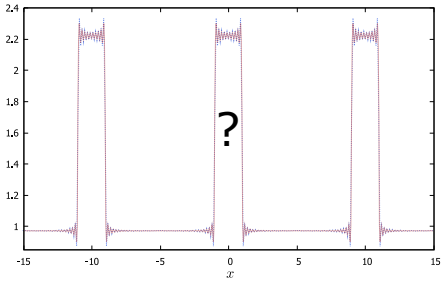
(b) Discrete, non-periodic spectrum $\Rightarrow$ infinitely large $[\epsilon]$



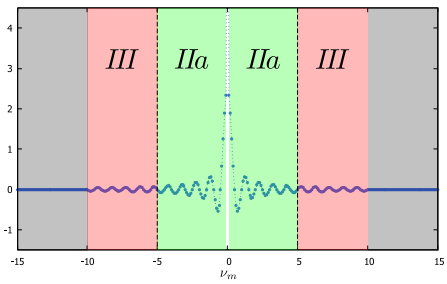
(c) Reconstructed, bandlimited $\epsilon^{(M_0)}(x)$: continuous and periodic



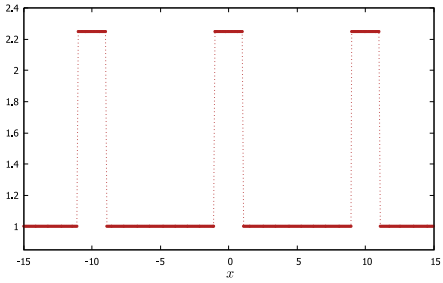
(d) Discrete, windowed spectrum $\Rightarrow$ infinitely large $[\epsilon]$



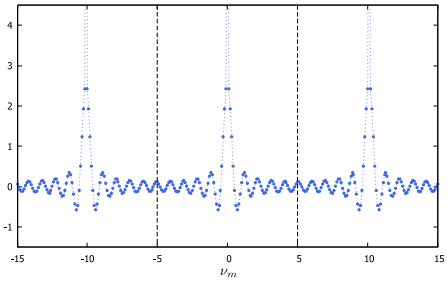
(e) Reconstruction of $\epsilon(x)$ undefined



(f) Erroneous windowed, discrete spectrum $\Rightarrow$ finite and square, but inconsistent $[\epsilon]$ used by the Std-RCWA



(g) Reconstructed, bandlimited $\epsilon^{(M_0)}(x_m)$: discrete and periodic



(h) Discrete, periodic spectrum from DFT $\Rightarrow$ finite, square and consistent $[\overset{\infty}{\epsilon}]$ used by the LIF-RCWA

Figure 3.3: Comparison of different bandwidth limiting schemes in position space (left-hand column) and Fourier spectrum (right-hand column).

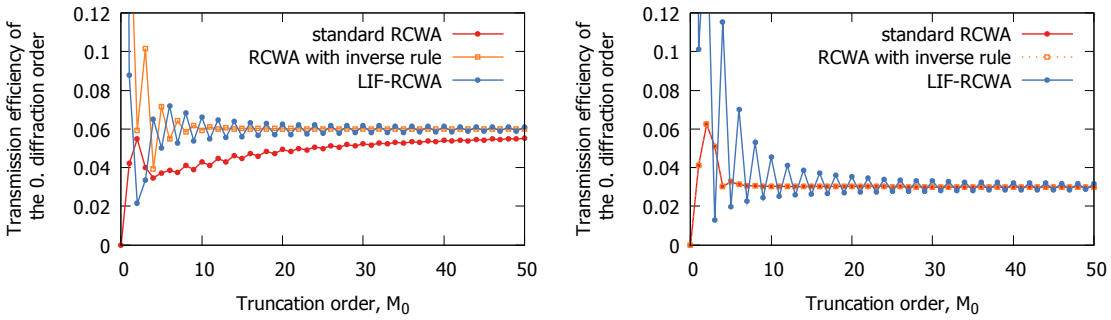
Ideally, a bandlimiting scheme should set all frequencies above a given cutoff frequency to zero. And this should also be reflected in the convolution matrix. The LIF approach calculates the Fourier coefficients of the permittivity with a discrete Fourier transform using the same number of sampling points that is later chosen for the truncation of the fields. This provides a periodic Fourier spectrum with the cyclic definitions for the electric field vector $\mathbf{S}$ and the convolution matrix $[\overset{\infty}{\epsilon}]$ of the permittivity as defined in Eqs. 3.9a and 3.9b. Thus, the bandwidth limiting step is already carried out before the truncation of the sums in Eq. 3.7c and all remaining spectral information are preserved without of any further approximations (cf. Fig. 3.3g). As such, a reasonable grating structure can be reconstructed from the convolution matrix (cf. Fig. 3.3h). This is not true for the standard RCWA approach.
On the other hand, if a grating structure has a bandwidth that is too large to practically obey the Nyquist–Shannon sampling theorem, cross talking between different spectral periods will occur, which causes aliasing effects. This issue becomes bigger in the presented LIF-approach, since the method requires the same number of sampling points in the spatial domain as it does in the frequency domain, while the latter defines the complexity of the RCWA algorithm and is therefore limited by computing resources. It is obvious, though, that for an increasing mode count the results of the two approaches equalize, since all discussed shortcomings arise from the demand of bandwidth limitation and mode truncation.

3.6 Properties of the circulant Toeplitz matrix

As discussed in Sec. 2.5, the accepted way to face the convergence problem of the RCWA in the TM case is the use of the inverse rule (cf. Eqs. 2.38a-2.38c), or more precisely the replacement of the Toeplitz matrix $[\epsilon]$ of the Fourier coefficients of $\epsilon(x)$ by the inverse Toeplitz matrix $[1/\epsilon]^{-1}$ of the Fourier coefficients of $1/\epsilon(x)$. The two matrices only match in approximation. This is because the convolution matrix of the standard approach only contains a subset of the permittivity's Fourier spectrum and therefore only approach each other for increasing truncation order. In the LIF-approach, on the other hand, the spectrum is periodic and the convolution matrix $[\overset{\infty}{\epsilon}]$ and its inverse contain a complete period and thus each hold the information of the entire discrete Fourier spectrum. As a consequence, the inverse rule is an exact operation, letting $[\overset{\infty}{\epsilon}]$ and $[\overset{\infty}{1/\epsilon}]^{-1}$ be two identical matrices (cf. Sec. 3.6.2). This fact, however, also makes the inverse rule obsolete for the LIF-approach. Another way of looking at this is revealed by the absence of any Gibbs phenomenon in the reconstructed permittivity in the LIF-RCWA (cf. Fig. 3.3g), which consequently spares the need for any sophisticated treatment. Of course, this is owed to the fact that the reconstruction is only performed at discrete positions. Nevertheless, it not only leads to a Gibbs-free reconstructed permittivity function, but also to an electric field without overshoots. Hence, the Gibbs phenomenon should not be the root of instability problems anymore, which led to the introduction of the inverse rule in the first place.

3.6.1 Example

In order to demonstrate the actual influence on the convergence behavior, the same metal grating as in Sec. 2.5.3 is analyzed (cf. Fig. 2.7) by the standard RCWA, the inverse-rule improved RCWA as well as the LIF-RCWA for TE- and TM-polarization.



(a) TM-Polaization

(b) TE-Polaization

Figure 3.4: Comparison of the convergence speeds of the standard RCWA (solid red dots), the LIF-approach (solid blue dots) and the standard approach with inverse rule applied (orange squares).

To begin with, in Figs. 3.4a and 3.4b it can be observed that the convergence speed of the standard approach (solid red dots) is considerably slower for TM-polarized light than for TE-polarized light. If the standard approach is, however, improved by the inverse rule (hollow red squares), the two convergence speeds suddenly match. For the LIF-approach, the convergence speeds for both polarization states are already similar. Compared with the standard approach, however, the LIF-approach (solid blue dots) clearly outperforms the unaltered standard approach in the TM case, while at the same time falls short compared to the TE- and the inverse-rule improved TM case.
This behavior can be attributed to the intrinsic weak convergence speed of the DFT approximation for functions with steep edges and thereby high frequencies. Since in the LIF-RCWA the number of spatial sampling points is linked to the number of modes, the Fourier approximation of the grating is usually weaker than in the standard RCWA. It also explains the higher amplitude oscillations of the LIF-RCWA, since a low sampling of the grating causes the reconstructed structures to be alternately either considerably smaller or larger than the original structures. This effect causes an inferior convergence speed of the LIF-approach in the TE case (cf. Fig. 3.4b). In case of TM-polarization (cf. Fig. 3.4a), on the other hand, the influence of the Gibbs phenomenon dominates, causing the LIF-RCWA to perform superior compared to the standard approach, since it does not suffer these instabilities.
Finally, the standard approach with the inverse rule applied even outperforms the LIF-RCWA in the TM case, because it not only eliminates the instabilities by the approximation of the inverse rule, but at the same time maintains the finer sampling of the standard approach. Nonetheless, the inverse rule does not fix the inconsistent bandwidth limitation, which prevents a reasonable grating reconstruction.

3.6.2 Proof of the inverse rule in the LIF-RCWA

For the mathematical proof the inverse rule in the LIF-RCWA a particular relation between circulant Toeplitz matrices and the DFT is used.
Any square $M \times M$ matrix $\mathbf{A}$ with M linearly independent eigenvectors $\mathbf{E}(\lambda_m)$ and the correspondent eigenvalues λ_m, $m = 0\,..\,M-1$ can be diagonalized and thus written as

$$\mathbf{A} = \mathbf{S}\mathbf{D}\mathbf{S}^{-1}. \tag{3.10}$$

In this equation the diagonal matrix $\mathbf{D}$ and the transformation matrix $\mathbf{S}$ are defined as $\mathbf{D} = \mathrm{diag}(\lambda_m)$ and $\mathbf{S} = \{\mathbf{E}(\lambda_0), \mathbf{E}(\lambda_1), ..., \mathbf{E}(\lambda_{M-1})\}$.

In the special case of a circulant Toeplitz matrix $[\overset{\infty}{c}]$ with $[\overset{\infty}{c}]_{j,k} = c_{\mathrm{wrap}(j-k,0,M)}$ and $j, k = 0\,..\,M-1$ the eigenvalues and eigenvectors are determined[42] by

$$\lambda_m = \sum_{j=0}^{N-1} c_j w_m^{-j} \tag{3.11a}$$

$$\text{and} \qquad \mathbf{E}(\lambda_m) = \frac{1}{\sqrt{M}} \left(1, w_m^1, w_m^2, ..., w_m^{N-1}\right)^T \tag{3.11b}$$

$$\text{with} \qquad w_m = e^{2\pi i \frac{m}{M}}. \tag{3.11c}$$

Using the definition of the Vandermonde matrix $\mathbf{F}_{M;m,j} = \exp\left(-2\pi i \frac{mj}{M}\right)$, Eqs. 3.11a and 3.11b can also be expressed in a more compact matrix-vector notation.

$$\lambda_m = \sum_{j=0}^{N-1} c_j e^{-2\pi i \frac{mj}{M}} \Rightarrow \lambda = \mathbf{F}_M \mathbf{c} \tag{3.12a}$$

$$\mathbf{S}_{m,j} = \frac{1}{\sqrt{M}} e^{2\pi i \frac{mj}{M}} \Rightarrow \mathbf{S} = \frac{1}{\sqrt{M}} \mathbf{F}_M^* \tag{3.12b}$$

Here, the interesting point is, that $\mathbf{F}_M$ can be identified as the transformation matrix of the discrete Fourier transform with $\mathbf{F}_M^*$ as its complex conjugate. Thus, considering $\mathbf{c}$ as a vector of real-space sampling points, $\mathbf{F}_M$ converts all entries into Fourier coefficients:

$$\tilde{\mathbf{c}} = \mathrm{DFT}(\mathbf{c}) = \frac{1}{M} \mathbf{F}_M \mathbf{c} \tag{3.13}$$

$\mathbf{S}$ is the matrix of eigenvectors and it is so of all circulant Toeplitz matrices. Furthermore, $\mathbf{S}$ is a unitary matrix with $\mathbf{S}^*\mathbf{S}^T = \mathbf{I}$ and therefore

$$\mathbf{S}^* = \frac{1}{\sqrt{M}}\mathbf{F}_M \tag{3.14a}$$
$$\mathbf{S} = (\mathbf{S}^*)^{-1} = \sqrt{M}\mathbf{F}_M^{-1}. \tag{3.14b}$$

Now, diagonalizing a circulant Toeplitz matrix leads to

$$[\overset{\infty}{c}] = \mathbf{S}\mathbf{D}\mathbf{S}^{-1} \tag{3.15a}$$
$$= \mathbf{S}\ \mathrm{diag}\left(\lambda_m\right)\mathbf{S}^* \tag{3.15b}$$
$$= \sqrt{M}\mathbf{F}_M^{-1}\,\mathrm{diag}\left(\mathbf{F}_M\mathbf{c}\right)\frac{1}{\sqrt{M}}\mathbf{F}_M \tag{3.15c}$$
$$= \mathbf{F}_M^{-1}\,\mathrm{diag}\left(\mathbf{F}_M\mathbf{c}\right)\mathbf{F}_M. \tag{3.15d}$$

And respectively the inverse of a circulant Toeplitz matrix can be diagonalized as

$$[\overset{\infty}{c}]^{-1} = (\mathbf{S}\ \mathrm{diag}\left(\lambda_m\right)\mathbf{S}^*)^{-1} \tag{3.16a}$$
$$= (\mathbf{S}^*)^{-1}\,\mathrm{diag}\left(\mathbf{F}_M\mathbf{c}\right)^{-1}\mathbf{S}^{-1} \tag{3.16b}$$
$$= \mathbf{S}\ \mathrm{diag}\left(\mathbf{F}_M\mathbf{c}\right)^{-1}\mathbf{S}^* \tag{3.16c}$$
$$= \sqrt{M}\mathbf{F}_M^{-1}\,\mathrm{diag}\left(\mathbf{F}_M\mathbf{c}\right)^{-1}\frac{1}{\sqrt{M}}\mathbf{F}_M \tag{3.16d}$$
$$= \mathbf{F}_M^{-1}\,\mathrm{diag}\left(\left(\mathbf{F}_M\mathbf{c}\right)^{-1}\right)\mathbf{F}_M. \tag{3.16e}$$

Replacing $\mathbf{c}$ by the Fourier coefficients $\tilde{\mathbf{c}}$, obtained by the discrete Fourier transformation, finally leads to the identity of the inverse rule:

$$[\overset{\infty}{\tilde{c}}]^{-1} = \mathbf{F}_M^{-1}\,\mathrm{diag}\left(\left(\mathbf{F}_M\tilde{\mathbf{c}}\right)^{-1}\right)\mathbf{F}_M \tag{3.17a}$$
$$= \mathbf{F}_M^{-1}\,\mathrm{diag}\left(\left(\frac{1}{M}\mathbf{F}_M\mathbf{F}_M\mathbf{c}\right)^{-1}\right)\mathbf{F}_M \tag{3.17b}$$
$$= \mathbf{F}_M^{-1}\,\mathrm{diag}\left(\frac{1}{M}\mathbf{F}_M\mathbf{F}_M\mathbf{c}^{-1}\right)\mathbf{F}_M \tag{3.17c}$$
$$= \mathbf{F}_M^{-1}\,\mathrm{diag}\left(\mathbf{F}_M\widetilde{(1./\mathbf{c})}\right)\mathbf{F}_M \tag{3.17d}$$
$$= \overset{\infty}{\widetilde{[1/c]}} \tag{3.17e}$$

The term $1./\mathbf{c}$ in Eq. 3.17d indicates an elementwise inversion of the vector elements in $\mathbf{c}$. The step from Eq. 3.17b to Eq. 3.17c is valid since $\frac{1}{M}\mathbf{F}_M\mathbf{F}_M$ is an anti-diagonal unit matrix. Hence, $\frac{1}{M}\mathbf{F}_M\mathbf{F}_M = \left(\frac{1}{M}\mathbf{F}_M\mathbf{F}_M\right)^{-1}$ and therefore also $\left(\frac{1}{M}\mathbf{F}_M\mathbf{F}_M\mathbf{c}\right)^{-1} = \frac{1}{M}\mathbf{F}_M\mathbf{F}_M\mathbf{c}^{-1}$.

The above derivation proves the exactness of the inverse rule for the Toeplitz matrix $[\tilde{c}]$, if the Fourier coefficients $\tilde{c}$ are calculated with discrete Fourier transform and $[\tilde{c}]$ is circular. In fact, the same considerations also apply, if the Fourier coefficients are calculated by a Fourier integral. However, in this case, the corresponding transformation matrix $\mathbf{F}_M$ would be of infinite size, which is consistent to the untruncated (but numerically impractical) scenario, where the inverse rule is already known to be exact - also for the standard RCWA. Furthermore, it should be noted, that the inverse rule does not hold for the truncated scenario in the standard RCWA, even if higher frequencies in the regions of class *III* (cf. Fig. 3.3f) are artificially set to zero and the coefficients of the electric field are truncated and not periodically repeated - although this would fix the truncation scheme of the standard RCWA.

3.7 Complexity analysis

In this section the complexity of the computation time and memory consumption of three different RCWA variants is compared using the big O notation. The comparison considers the calculation of the diffraction coefficients of a two-dimensional diffraction problem, which refers to the diffraction at a one-dimensional, single-layered grating in classical mounting, and a three-dimensional diffraction problem, which refers to a two-dimensional, single-layered grating at arbitrary incidence. The first method is the standard RCWA (Std-RCWA) as it was published by Moharam and Gaylord[91], the second one is the localized input field RCWA of this work (LIF-RCWA)[6] and the third one is the finite-beam RCWA (FB-RCWA) of Wu and Gaylord[133].

3.7.1 Time complexity comparison

The runtime of each algorithm is described by $T(M_0)$, which is a function of the truncation order M_0. Its complexity for the standard RCWA is primarily determined by the eigenvalue problem (EV), which needs to be solved for the matrix $\mathbf{\Omega}_1$ (cf. Eq. 2.16). For a one-dimensional grating the dimension of the matrix is $(2M_0+1)$[88], for a two-dimensional grating it is $2(2M_0+1)^2$ (cf. Eq. 2.16) (assuming an equal sampling in both lateral dimensions). This leads to the following time complexity:

$$T^{\mathrm{1D}}_{\mathrm{EV}}(M_0) \in O(M_0^3) \;\Rightarrow\; T^{\mathrm{2D}}_{\text{Std-RCWA}}(M_0) \in O(M_0^3), \tag{3.18a}$$

$$T^{\mathrm{2D}}_{\mathrm{EV}}(M_0) \in O(M_0^6) \;\Rightarrow\; T^{\mathrm{3D}}_{\text{Std-RCWA}}(M_0) \in O(M_0^6). \tag{3.18b}$$

The additional computation effort in the LIF-RCWA is confined to the ASD, which is basically a fast Fourier transform:

$$T^{\mathrm{1D}}_{\mathrm{FFT}}(M_0) \in O(M_0 \log(M_0)) \;\Rightarrow\; T^{\mathrm{1D}}_{\mathrm{ASD}}(M_0) \in O(M_0 \log(M_0)) \tag{3.19a}$$

$$T^{\mathrm{2D}}_{\mathrm{FFT}}(M_0) \in O(2M_0^2 \log(M_0^2)) \;\Rightarrow\; T^{\mathrm{2D}}_{\mathrm{ASD}}(M_0) \in O(2M_0^2 \log(M_0^2)) \tag{3.19b}$$

With the simplification rules of the big O notation, the complexity of the LIF-RCWA can easily be determined. However, an additional constant, say c, needs to be considered, since the LIF-RCWA shows a slower convergence rate due to the lower sampling of the permittivity:

$$T^{\text{2D}}_{\text{LIF-RCWA}}(M_0) = T^{\text{1D}}_{\text{ASD}}(M_0) + T^{\text{2D}}_{\text{Std-RCWA}}(cM_0) \in O(M_0^3) \tag{3.20a}$$
$$T^{\text{3D}}_{\text{LIF-RCWA}}(M_0) = T^{\text{2D}}_{\text{ASD}}(M_0) + T^{\text{3D}}_{\text{Std-RCWA}}(cM_0) \in O(M_0^6) \tag{3.20b}$$

Suggesting that the FB-RCWA uses the same truncation order for the ASD as for the calculation of the permittivity coefficients like the LIF-RCWA does, the FB-RCWA shows a considerably higher complexity as it calculates a full standard RCWA for every decomposed plane wave:

$$T^{\text{2D}}_{\text{FB-RCWA}}(M_0) = T^{\text{1D}}_{\text{ASD}}(M_0) + M_0 \cdot T^{\text{2D}}_{\text{Std-RCWA}}(M_0) \in O(M_0^4) \tag{3.21a}$$
$$T^{\text{3D}}_{\text{FB-RCWA}}(M_0) = T^{\text{2D}}_{\text{ASD}}(M_0) + M_0^2 \cdot T^{\text{3D}}_{\text{Std-RCWA}}(M_0) \in O(M_0^8) \tag{3.21b}$$

In summary, the complexity analysis reveals that the LIF-RCWA has the same time complexity as the standard RCWA, but a significantly lower time complexity than the FB-RCWA, which is also able to simulate the diffraction of localized beams.

3.7.2 Space complexity comparison

The storage requirements of the different algorithms shall be described by $S(M_0)$ and its complexity can be derived similarly to the time complexity in the previous section. The space complexity depends on the data structure with the largest size in terms of magnitude, which is again the matrix $\mathbf{\Omega}_1$. For a two-dimensional grating problem Eq. 2.16 shows that the calculation of $\mathbf{\Omega}_1$ requires two additional matrices $\mathbf{M}_1$ and $\mathbf{M}_2$ of the same size, which each consists of four submatrices due to the two expansion dimensions. Each submatrix has the size of $(2M_0+1)^2 \times (2M_0+1)^2$ elements - again assuming an equal sampling in both dimensions. For a one-dimensional grating problem in classical mounting the size of matrix $\mathbf{\Omega}_1$ reduces to $(2M_0+1) \times (2M_0+1)$ elements, since Fourier expansion is only performed in one dimension and due to the separation in two different types of polarization, which only requires a single lateral component to be considered at the same time. This leads to a space complexity of

$$S^{\text{1D}}_{\mathbf{\Omega}_1}(M_0) \in O(M_0^2) \;\Rightarrow\; S^{\text{2D}}_{\text{Std-RCWA}}(M_0) \in O(M_0^2), \tag{3.22a}$$
$$S^{\text{2D}}_{\mathbf{\Omega}_1}(M_0) \in O(M_0^4) \;\Rightarrow\; S^{\text{3D}}_{\text{Std-RCWA}}(M_0) \in O(M_0^4). \tag{3.22b}$$

If the Angular Spectrum Method uses a discrete Fourier transform, the incident field is described by K and K^2 sampling points for the one- and two-dimensional description respectively. The discretized incident field is then transformed into $2M_0+1$ and $(2M_0+1)^2$ Fourier modes.

Since the LIF-RCWA uses the same sampling rate in the spatial and frequency domain, the number of sampling points is also $2M_0+1$ and $(2M_0+1)^2$. Usually K is chosen to be larger than $2M_0+1$. The complexity of the ASD can then be derived as

$$S^{\mathrm{1D}}_{ASD,FB-RCWA}(K) \quad \in O(K), \tag{3.23a}$$
$$S^{\mathrm{2D}}_{ASD,FB-RCWA}(K^2) \in O(K^2), \tag{3.23b}$$
$$S^{\mathrm{1D}}_{ASD,LIF-RCWA}(M_0) \in O(M_0), \tag{3.23c}$$
$$S^{\mathrm{2D}}_{ASD,LIF-RCWA}(M_0^2) \in O(M_0^2). \tag{3.23d}$$

The LIF-RCWA only performs one ASD and one RCWA calculation to obtain the $2M_0+1$ or $(2M_0+1)^2$ reflection and transmission coefficients. The FB-RCWA on the other hand needs to store at least these $2\cdot 2M_0+1$ or $2\cdot(2M_0+1)^2$ coefficients of K or K^2 separate RCWA calculations in order to superpose the results at the end. The intermediate $\mathbf{\Omega}_1$ matrices however can be dropped. This, finally, results in space complexities of

$$S^{\mathrm{2D}}_{FB-RCWA}(K) \in O(M_0^2 + M_0K), \tag{3.24a}$$
$$S^{\mathrm{3D}}_{FB-RCWA}(K^2) \in O(M_0^4 + M_0^2K^2), \tag{3.24b}$$
$$S^{\mathrm{2D}}_{LIF-RCWA}(M_0) \in O(M_0^2), \tag{3.24c}$$
$$S^{\mathrm{3D}}_{LIF-RCWA}(M_0^2) \in O(M_0^4). \tag{3.24d}$$

As a result, the space complexity of the LIF-RCWA is equal to that of the standard RCWA. The space complexity of the FB-RCWA strongly depends on the sampling of the incident field and might in practice be higher but with the advantage of a higher resolution.

3.8 Diffraction order efficiencies

Similar to the efficiency considerations of the standard RCWA (cf. Sec. 2.9), the relative power per diffraction order needs to be set in relation to the total power of the incident field. In the LIF-approach, this refers to the total power of multiple input modes rather than a single one:

$$\text{Std-RCWA:} \qquad \eta^T_{m,n} := \frac{[\langle \mathbf{S}_t \rangle d\mathbf{A}_t]_{m,n}}{\langle \mathbf{S}_i \rangle d\mathbf{A}_i} \tag{3.25a}$$

$$\text{LIF-RCWA:} \qquad \eta^T_{m,n} := \frac{[\langle \mathbf{S}_t \rangle d\mathbf{A}_t]_{m,n}}{\sum [\langle \mathbf{S}_i \rangle d\mathbf{A}_i]_{m,n}} \tag{3.25b}$$

As described in Eq. 2.64b $\langle \mathbf{S} \rangle d\mathbf{A}$ denotes the average power as a product of the time averaged energy flux (Poynting vector $\mathbf{S}$) and a surface vector $d\mathbf{A}$ of the irradiated area.

With few simple steps of calculation, analog to those performed in Sec. 2.9, the diffraction order efficiencies $\eta^T_{m,n}$ and $\eta^R_{m,n}$ for transmission and reflection in the LIF-approach can be derived. In addition, an expression for the relative power $\eta^I_{m,n}$ of an input mode can also be obtained:

$$L_0 = \frac{1}{\mu_I} \sum_{m,n} |L_{m,n}|^2 Re\left(\gamma_{I;m,n}\right), \tag{3.26a}$$

$$\eta^T_{m,n} = \frac{1}{\mu_{III}} \frac{|T_{m,n}|^2}{L_0} Re\left(\gamma_{III;m,n}\right), \tag{3.26b}$$

$$\eta^R_{m,n} = \frac{1}{\mu_I} \frac{|R_{m,n}|^2}{L_0} Re\left(\gamma_{I;m,n}\right), \tag{3.26c}$$

$$\eta^I_{m,n} = \frac{1}{\mu_I} \frac{|L_{m,n}|^2}{L_0} Re\left(-\gamma_{I;m,n}\right) \tag{3.26d}$$

For non-absorbing gratings, consistency with the law of energy conservation should still preserved by $1 - \sum \eta^T_{m,n} - \sum \eta^R_{m,n} = 1 - \mathcal{R} - \mathcal{T} = 0$.

3.9 Modeling of a localized input fields

The most important novelty of the LIF-approach is the ability to natively solve the diffraction problem not only for a single, infinite plane wave, but also for a localized incident field. Furthermore, it also allows various forms of polarization. Both properties are achieved by individually defining the amplitudes of the discrete function that describes the incident light field in the pupil. They are given in the vector $\mathbf{L}$ of Eq. 3.2, which can be broken down to a scalar amplitude term $\mathbf{A}_\perp$ and a polarization term $\mathbf{\Psi}\left(\psi\right)$ independently for each mode (m, n):

$$\mathbf{L}_{\perp;m,n} = \mathbf{A}_{m,n} \mathbf{\Psi}_{\perp;m,n}\left(\psi\right) \tag{3.27}$$

3.9.1 Angular spectrum decomposition

The angular spectrum decomposition (ASD) was already introduced as an integral part of the Angular Spectrum Method in Sec. 1.1.6 for the propagation of plane waves in a homogeneous medium. Although the main focus of the RCWA are the propagation through inhomogeneous media, the decomposition into plane waves can also be used to create a suitable input field for the LIF-approach. This is possible since in the LIF-RCWA just like in the standard RCWA the incident region is assumed to be homogeneous. As proposed at the beginning of this chapter, a localized incident field can be modeled by a superposition of multiple plane waves, which is described by the Fourier series in Eq. 3.2. In fact this is analogous to the

modeling of the grating permittivity in Eq. 2.1. Likewise, the Fourier coefficients of an incident field can either be calculated by solving the correspondent Fourier integral (cf. Eq. 2.2) or by calculating the discrete Fourier transform. However, since the earlier discussion in Sec. 3.4 about the prerequisites of the LIF-RCWA resulted in the demand for a discrete electric fields with a bandwidth limitation equal to the grating, only the discrete Fourier transform, or in this case also called Angular Spectrum Decomposition, is the suited method for the LIF-approach and is given by

$$A_{m,n} = \frac{A_0}{MN} \sum_{p,q} A\left(x_p, y_q, 0\right) e^{-2\pi i \left(\frac{m}{P_x} x_p + \frac{n}{P_y} y_q\right)}. \tag{3.28}$$

$A\left(x_p, y_q, 0\right)$ describes the scalar amplitude of the electric field distribution at a single sampling point $(x_p, y_q, 0)$ in the x/y plane at the grating interface at $z = 0$. In the following, this plane is referred to as the plane of origin. Due to the enforced periodicity and the sampling of position space, the ASD in Eq. 3.28 is adjusted to match the overall number of sampling points $M \times N$ as well as the discrete frequencies ($\nu_m = m/P_x$ and $\nu_n = n/P_y$) of the discrete Fourier decomposition of the grating. As a consequence, also the light source is now periodic and the desired locality of the input field always refers to a position in every period. Hence, simulating the focusing of a light beam onto a grating structure requires the definition of a grating *superperiod*, which itself may contain finer substructures (cf. Fig. 3.10). and might require a higher truncation limit. In order to distinguish between the two periods in the following, they are referred to as P' for the period of substructures and P to denote the period of the superstructure that also determines the frequencies of the discrete Fourier transform ($\delta\nu = 1/P$). In order to avoid crosstalk between neighboring superperiods, absorbing boundaries or the concept of perfectly matched layers[114] can be used.
Finally, A_0 in Eq. 3.28 is a normalization factor to assure that the incident power over one grating period is always one:

$$\sum_{m,n} |A_{m,n}|^2 = 1. \tag{3.29}$$

Thus, with definition 3.28 any sufficiently sampled light field can be used as an input for the rigorous analysis in the LIF-RCWA - if only its distribution in the plane of origin is known.

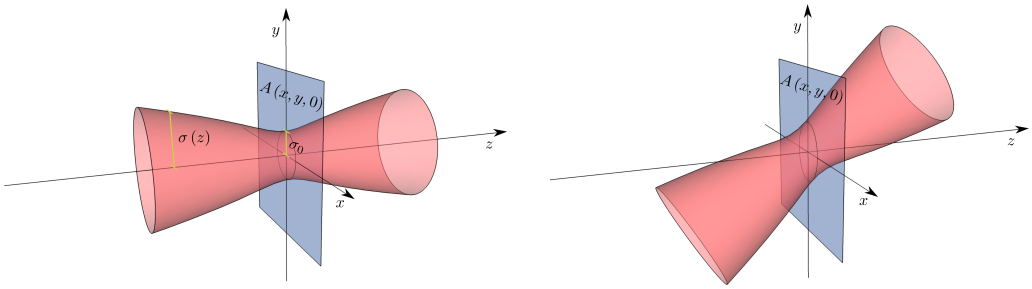
Example: paraxial Gaussian beam model

For a paraxial Gaussian beam at normal incidence and with a beam waist at $z = 0$ (cf. Fig. 3.5a), the field distribution in the plane of origin can be described by the Gaussian function

$$A_G^{2D}\left(x, y, 0\right) = \exp\left(-2\frac{x^2 + y^2}{\sigma_0^2}\right). \tag{3.30}$$

Here, σ_0 denotes the size of the beam's waist.

According to the ASD (cf. Eq. 3.28) Eq. 3.30 already provides sufficient information for the LIF-RCWA to rigorously propagate a Gaussian beam. It should be noted, however, that the described Gaussian beam model does not rigorously satisfy the wave equation but only its paraxial approximation[138].



(a) Normal Gaussian beam (b) Tilted Gaussian beam

Figure 3.5: Determination of the field distribution of a Gaussian beam in the plane of origin

In order to propagate a tilted Gaussian beam, the ASD requires a different slice of the beam's field distribution (cf. Fig. 3.5b). Therefore, the beam must also be known in the near surrounding of its center. This can be approximated by paraxial propagation. Therefore, the Fresnel-transformation

$$A\left(\mathbf{r}_\perp, z\right) = -\frac{i}{\lambda z} e^{izn k_0} \iint A\left(\mathbf{r}'_\perp, 0\right) e^{i\frac{\pi}{\lambda z}\left(\mathbf{r}_\perp - \mathbf{r}'_\perp\right)^2} d^2\mathbf{r}'_\perp, \tag{3.31}$$

which is a paraxial approximation of the Frauenhofer diffraction integral, can be solved with the definition of A_G^{2D} (cf. Eq. 3.30) leading to[15]

$$A_G^{3D}\left(x, y, z\right) = \left(\frac{\sigma\left(z\right)}{w\left(z\right)}\right)^2 p\left(z\right) e^{-\frac{x^2+y^2}{w(z)^2}} \tag{3.32a}$$

$$\text{with} \qquad \sigma\left(z\right) = \sigma_0 \sqrt{1 + \frac{\lambda^2 z^2}{\pi^2 \sigma_0^4}}, \tag{3.32b}$$

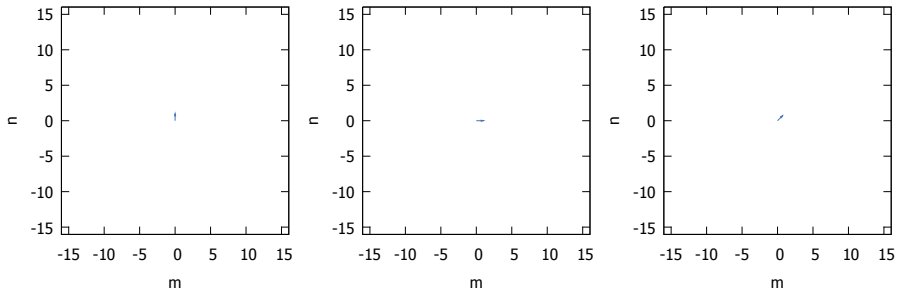
$$w\left(z\right) = \sqrt{\sigma\left(z\right) + i\frac{\lambda}{\pi} z} \qquad \text{and} \tag{3.32c}$$

$$p\left(z\right) = e^{ink_0 z}. \tag{3.32d}$$

Now, choosing an arbitrary tilted or shifted plane in $A_G^{3D}\left(x, y, z\right)$ as the field distribution in the plane of origin - or more precisely as $A_G^{2D}\left(x, y, 0\right)$ - allows its rigorous propagation and therewith the simulation of a tilted or shifted Gaussian beam in the LIF-RCWA.

3.9.2 Polarization

Single mode polarization



(a) $\mathbf{\Psi}_\perp(\psi = 90 \text{ deg})$ ⇒ TE Polarization (b) $\mathbf{\Psi}_\perp(\psi = 0 \text{ deg})$ ⇒ TM Polarization (c) $\mathbf{\Psi}_\perp(\psi = 45 \text{ deg})$

Figure 3.6: Illustration of the electric field polarization of a single mode input at different polarization angles ψ. The arrows indicate the tangential direction of the electric field in the x/y-plane.

In the standard RCWA, the input field is characterized by only a single mode and its polarization can be fully described by the polarization angle ψ between the electric field vector and the plane of incidence (cf. Fig. 2.1 and Eq. 1.26):

$$\mathbf{\Psi}(\psi) = \sin(\psi)\,\mathbf{e}_N + \cos(\psi)\,\mathbf{e}_T \tag{3.33a}$$

$$= \sin(\psi)\left[\frac{1}{s_\perp}\begin{pmatrix}-s_y\\ s_x\\ 0\end{pmatrix}\right] + \cos(\psi)\left[\frac{1}{s_\perp}\begin{pmatrix}s_x s_z\\ s_y s_z\\ -s_\perp^2\end{pmatrix}\right] \tag{3.33b}$$

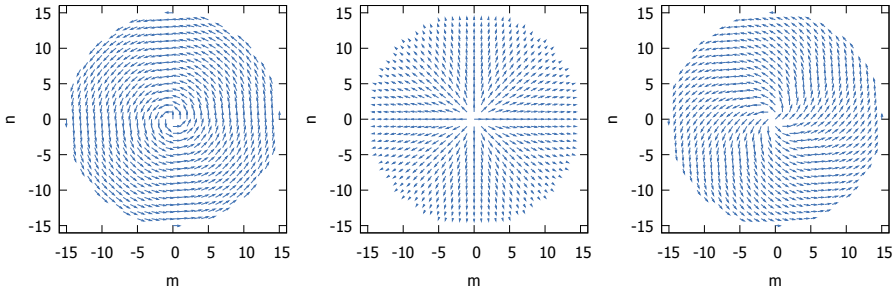
With the unit vectors $\mathbf{e}_N$ and $\mathbf{e}_T$ normal and tangential to the incident plane as defined in Eqs. 1.27a and 1.27b, $\mathbf{\Psi}(\psi)$ describes a superposed polarization state for a single plane wave. In classical mounting (cf. Fig. 1.3a) the basic polarization states at $\psi = 90$ deg and $\psi = 0$ deg correspond to TE- and TM-polarization respectively (cf. Fig. 3.6a and 3.6b). An intermediate state at $\psi = 45$ deg is shown in Fig. 3.6c.

Multi mode polarization

Definition 3.33b can easily be generalized for a multi mode field:

$$\mathbf{\Psi}(\psi)_{\perp;m,n} = \sin(\psi)\underbrace{\left[\frac{1}{s_{\perp;m,n}}\begin{pmatrix}-s_{y;m}\\ s_{x;n}\\ 0\end{pmatrix}\right]}_{\mathbf{e}_{N;m,n}} + \cos(\psi)\underbrace{\left[\frac{1}{s_{\perp;m,n}}\begin{pmatrix}s_{x;m}s_{z;m,n}\\ s_{y;n}s_{z;m,n}\\ -s_{\perp;m,n}^2\end{pmatrix}\right]}_{\mathbf{e}_{T;m,n}} \tag{3.34}$$

However, as illustrated in Figs. 3.7a - 3.7c, this does not lead to linear polarization, as one might expect. Instead, the angles $\psi = 90$ deg and $\psi = 0$ deg correspond to azimuthal and radial polarization. An intermediate state at $\psi = 45$ deg is again shown in Fig 3.7c.



(a) $\mathbf{\Psi}_\perp(\psi = 90 \text{ deg})$ ⇒ azimuthal polarization

(b) $\mathbf{\Psi}_\perp(\psi = 0 \text{ deg})$ ⇒ radial polarization

(c) $\mathbf{\Psi}_\perp(\psi = 45 \text{ deg})$ ⇒ mixed polarization

Figure 3.7: Examples of the electric field polarization of a multi-mode input at different polarization angles ψ. For every discrete point (m, n) in k-space an arrow indicates the direction of the tangential polarization in the x/y-plane. Shorter arrows refer to a larger normal(z)-component.

Linear polarization

True linear polarization was derived by Mansuripur[80] for $\psi = 0$ deg and generalized by Fragello[37]. A more compact and intuitive form is given by Brenner[16] and therefore also utilized in the present approach.

The intuitive way of defining linear polarization would be in a two-dimensional Cartesian coordinate system like

$$\mathbf{\Psi}'(\psi)_{m,n} = \sin(\psi)\,\mathbf{e}_y + \cos(\psi)\,\mathbf{e}_x \tag{3.35}$$

or equally in cylindric coordinates like

$$\mathbf{\Psi}'(\psi) = \underbrace{\left(\mathbf{\Psi}'(\psi)\,\mathbf{e}'_N\right)}_{\mathbf{\Psi}'_N}\mathbf{e}'_N + \underbrace{\left(\mathbf{\Psi}'(\psi)\,\mathbf{e}'_T\right)}_{\mathbf{\Psi}'_T}\mathbf{e}'_T \tag{3.36}$$

with

$$\mathbf{e}'_N = \frac{1}{s_{\perp;m,n}}\begin{pmatrix} -s_{y;m} \\ s_{x;n} \end{pmatrix}, \tag{3.37a}$$

$$\mathbf{e}'_T = \frac{1}{s_{\perp;m,n}}\begin{pmatrix} s_{x;m} \\ s_{y;n} \end{pmatrix}. \tag{3.37b}$$

However, this is only true for a plane wave with normal incidence ($\mathbf{s}_{i;m,n} = \mathbf{e}_z$) like the 0th mode of the ASD, because the electric field vector $\mathbf{L}_{m,n}$ of every (m,n)th plane wave is required to be normal to its corresponding wave vector $\mathbf{k}_{i;m,n} = n_I k_0 \mathbf{s}_{i;m,n}$.

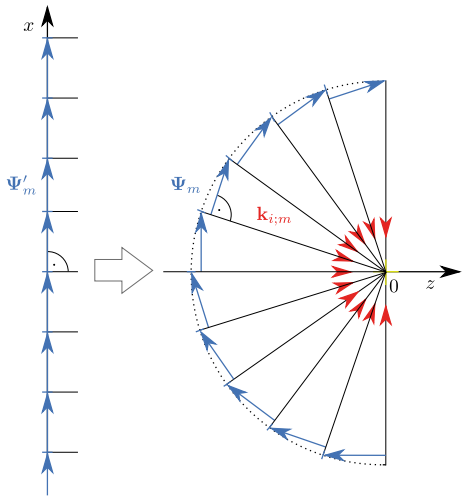
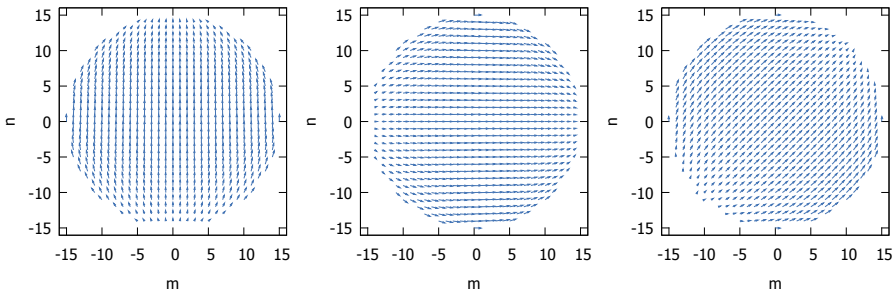


Figure 3.8: Transformation of the polarization from input plane to pupil sphere in a two-dimensional view

In fact, k-space alias the pupil of an optical system is filled with wave vectors pointing from all directions in the left hemisphere to the point of origin (cf. Fig. 3.8 and Sec. 3.10.3). Hence, the polarization vectors need to be mapped on to the surface of the pupil sphere. This can be achieved by a simple replacement of the cylindric coordinates $\mathbf{e}'_{N;m,n}$ and $\mathbf{e}'_{T;m,n}$ by the spherical coordinates $\mathbf{e}_{N;m,n}$ and $\mathbf{e}_{T;m,n}$ for every polarization vector:

$$\begin{aligned} \mathbf{\Psi}(\psi)_{m,n} &= \mathbf{\Psi}'_N \mathbf{e}_{N;m,n} + \mathbf{\Psi}'_T \mathbf{e}_{T;m,n} \\ &= \sin(\psi) \left[\frac{1}{s_\perp^2} \begin{pmatrix} s_x s_y (s_z - 1) \\ s_y^2 s_z + s_x^2 \\ -s_y s_\perp^2 \end{pmatrix} \right]_{m,n} + \cos(\psi) \left[\frac{1}{s_\perp^2} \begin{pmatrix} s_x^2 s_z + s_y^2 \\ s_x s_y (s_z - 1) \\ -s_x s_\perp^2 \end{pmatrix} \right]_{m,n} \end{aligned} \tag{3.38}$$

The result can can be verified in Figs. 3.9a - 3.9c.



(a) $\mathbf{\Psi}_\perp(\psi = 90 \text{ deg})$ ⇒ y-polarization

(b) $\mathbf{\Psi}_\perp(\psi = 0 \text{ deg})$ ⇒ x-polarization

(c) $\mathbf{\Psi}_\perp(\psi = 45 \text{ deg})$ ⇒ slanted polarisation

Figure 3.9: Illustration of linear electric field polarization of a multi-mode input at different polarization angles ψ.

3.10 Examples

In the following section different examples demonstrate the potential of the LIF-RCWA.

3.10.1 Phase Ronchi

In the first example, the response of a grating structure is analyzed, which is illuminated by a localized beam. The grating is a phase Ronchi with a (sub-)period of $P'_x = 4\,\mu m$ and consists of two refractive indices $n_{II,a} = 1.5$ and $n_{II,b} = 1$ at a duty cycle of 0.4 (cf. Fig. 3.10).

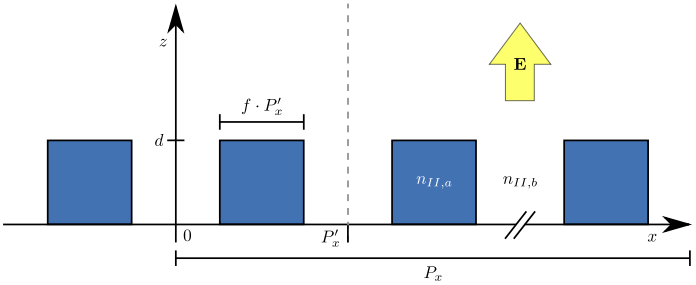


Figure 3.10: Sketch of a Ronchi grating

The thickness of the grating is $d = 1\,\mu m$ and the refractive index of the surrounding superstrate and the substrate is $n_I = n_{III} = 1$. The grating is illuminated by a Gaussian beam with a waist of $\sigma_0 = 8\,\mu m$ (cf. Eq. 3.30) at a wavelength of $\lambda = 1\,\mu m$ and with linear polarization at $\psi = 0$ deg. In order to allow the localized beam to cover multiple periods of the grating, a superperiod of $P_x = 16 \times P'_x$ is chosen. The Fourier coefficients of the permittivity and the Gaussian beam are calculated with discrete Fourier transform and $M = 2M_0 + 1$ sampling points. Hence, higher harmonics greater than $M_0 = 122$ are neglected.

Fig. 3.11 shows the intensity distribution of the electric near field along the propagation direction (z): As the Gaussian beam hits the grating it divides into two distinct diffraction orders with a diffraction angle of $\theta_m = \arcsin\left(m\lambda / \left(n_I P'_x\right)\right) < 90$ deg $\Rightarrow \theta_{\pm 1} = \pm 14.478$ deg. Higher orders are not visible. The correctness of the diffraction angles can be verified more easily in Fig. 3.12a, which shows a larger section of the simulation area and in particular the crossing of two diffraction orders of neighboring superperiods. It also reminds of the fact, that the light source is always periodic in the LIF-RCWA. Fig. 3.12b shows the attempt to emulate the same setup with the standard RCWA in combination with an aperture. Here, a plane wave is masked by implanting an imaginary epsilon of $\mathrm{Im}(\epsilon_{II,mask}) = 4$ inside the grating layer, leaving only a slit per superperiod for the light to pass. The slit was $0.2P_x = 12.8\,\mu m$ in size. Again, the correct diffraction angles can be verified by reading off the z-position of the beams' crossing. In both figures also the overall reflection and transmission efficiencies are presented. While in the LIF-RCWA they sum up to 100 % according to the law of energy

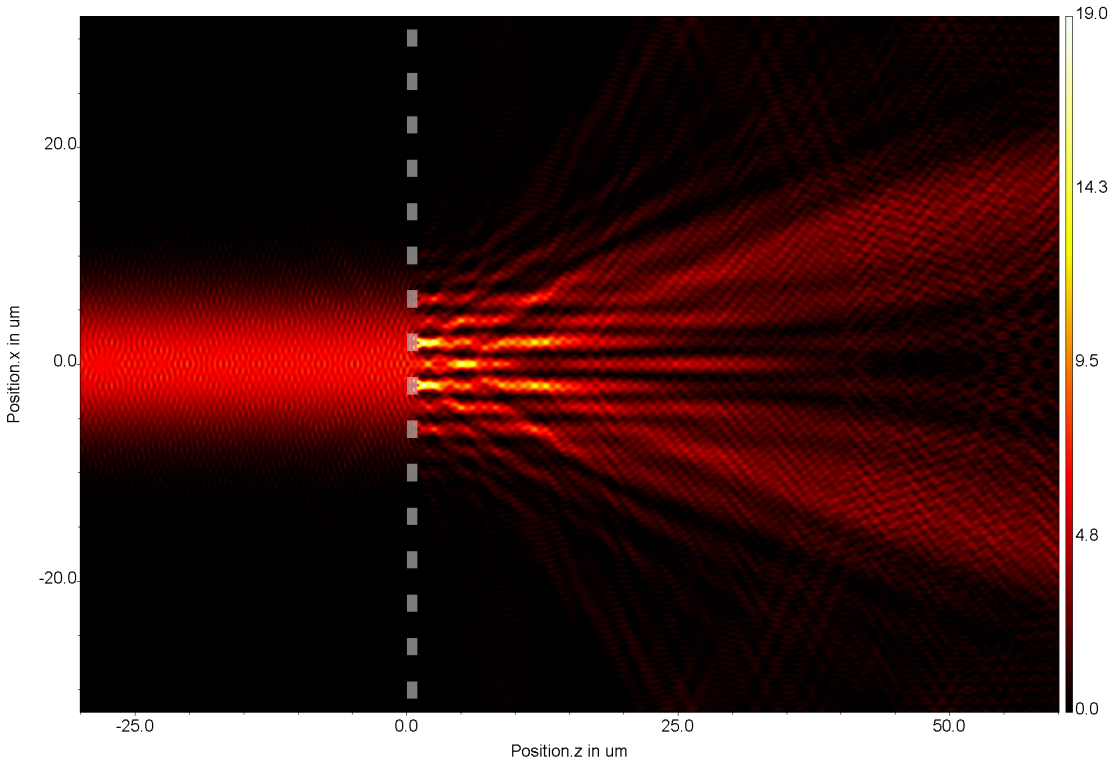
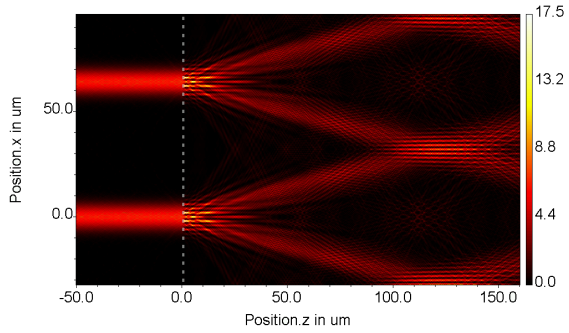


Figure 3.11: Electric near field intensity of a linearly polarized Gaussian beam with a waist size of $\sigma_0 = 8\,\mu m$ and wavelength of $\lambda = 1\,\mu m$, which is illuminating a Ronchi ruling (highlighted) with a period of $P'_x = 4\,\mu m$, a duty cycle of 0.4, refractive indices $n_{II,a} = 1.5$, $n_{II,b} = 1$, $n_I = n_{III} = 1$ and a thickness of $d = 1\,\mu m$.

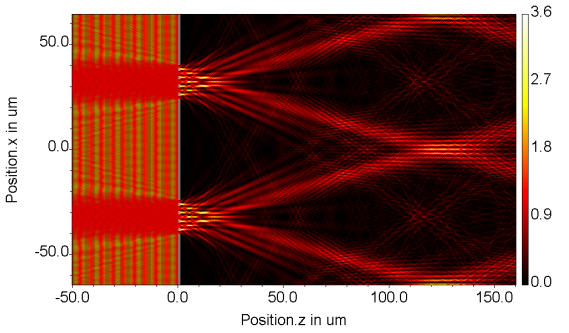
conservation, they do not in the standard approach. The difference marks the amount of dissipative loss at the mask, also known as absorption. The two results show great analogy in respect of the diffraction angles. However, due to the different natures of their beam profiles, there are considerable differences in the pattern of the near field and especially more scattering artifacts in the standard approach.
Following the definition for a tilted Gaussian beam in Sec. 3.9.1, the LIF-RCWA can also simulate beams of slanted incident. Fig. 3.13 shows the same setup with an incidence angle of 10 *deg*.

For Fig. 3.14 the same grating is illuminated with a Gaussian beam of increasing beam waist. The two graphs show the corresponding transmission efficiencies for the 0th (hollow red squares) and ±1st (solid blue circles) diffraction order. On the left end of the x-axis, one can observe that two grating periods are already more than sufficient to produce distinct diffraction orders. With increasing beam size, the two graphs converge towards the plane wave solutions (dashed lines), which were obtained by the standard RCWA. The limit of the transmission efficiency of the zeroth diffraction order is 2.7639 %, whereas it is 38.8703 % for the transmission efficiency of the ±1th diffraction order. With $P'_x/\lambda \ll 10$ and $d/\lambda = 1$, a rigorous analysis, which respects the full vectorial nature of light, is required for the calculation of the efficiencies. And hence, the use of simplified scalar diffraction theories is not legitimate anymore[56].



(a) LIF-RCWA simulation: Gaussian beam illumination with $\sigma_0 = 8\,\mu m$

Efficiencies: $T = 95.711\,\%, R = 4.289\,\%$



(b) Standard RCWA simulation: Masked plane wave illumination with an aperture of $12.8\,\mu m$

Efficiencies: $T = 22.946\,\%, R = 16.940\,\%$

Figure 3.12: Comparison between localized beams in the LIF-RCWA (a) and the standard RCWA (b).

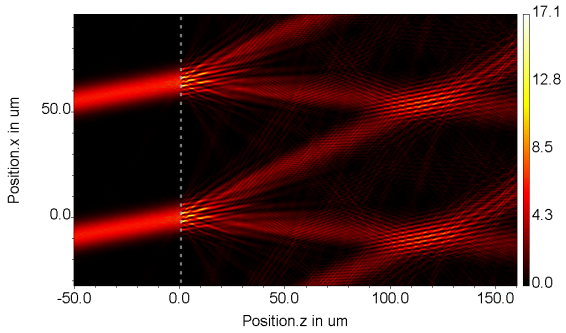


Figure 3.13: Slanted Gaussian beam with an incident angle of $\theta = 10$ deg

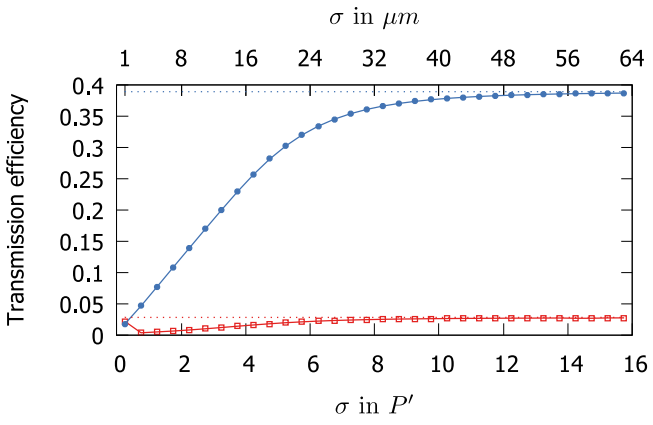
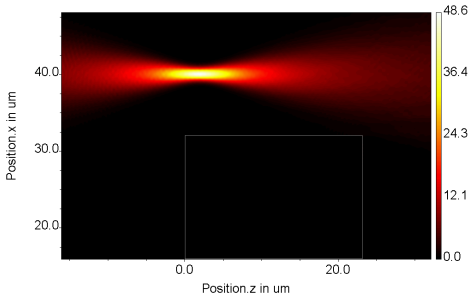


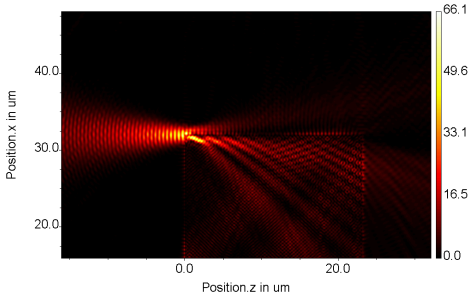
Figure 3.14: Transmission efficiency of the 0th (hollow red squares) and ±1st (solid blue circles) diffraction order of a Gaussian beam illumination with increasing waist size σ_0. The transmission efficiencies of the plane wave illumination (dashed lines) are marking the convergence limit

3.10.2 Focusing into an inhomogeneous medium

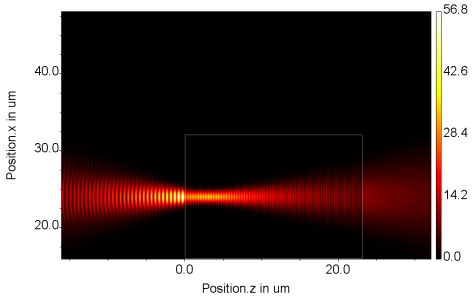
In the previous example one could have used the standard RCWA with a masked plane wave to roughly emulate a localized beam as done in Fig. 3.12b. However, if the optical setup requires a focused, converging-diverging Gaussian beam, masking alone is not sufficient anymore. In the following setup a converging Gaussian beam is moved over the edge of a structure. This is a scenario typically occurring in the reading process of optical media, like CDs, DVDs and Blu-rays.
In the example, the waist of the Gaussian beam has a size of $\sigma_0 = 1\,\mu m$ and is positioned at $z = +2\,\mu m$. Its wavelength is $\lambda = 1\,\mu m$. The illuminated object has a refractive index of $n_{a;II} = 1.5$, while the surrounding medium has an index of $n_I = n_{b;II} = n_{III} = 1$. The simulation was performed with a total of $M = 285$ modes. Figs. 3.15a - 3.15c show the passing of the light beam through the structure for various x-positions of the beam. While the Gaussian beam seems undisturbed in Fig. 3.15a as it only experiences a homogeneous medium, the beam splits at the corner of the structure (cf. Fig. 3.15b) resulting in a significant decrease in transmission efficiency (cf. Fig. 3.16). Inside the denser medium the beam's waist appears compressed (cf. Fig. 3.15c). The striped patterns in Fig. 3.15b and 3.15c arise from the interference with the reflected beam at the front and rear interface of the structure.
Fig. 3.16 shows the transmission efficiency as a function of the spot position.



(a)



(b)



(c)

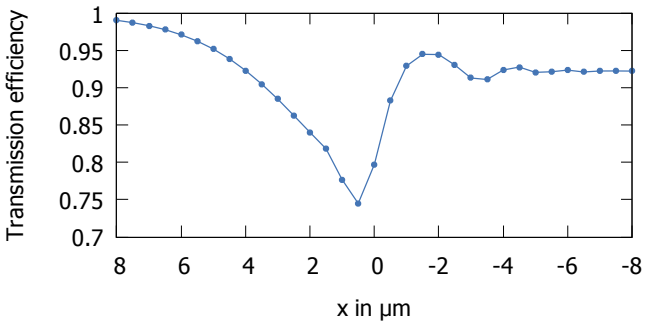
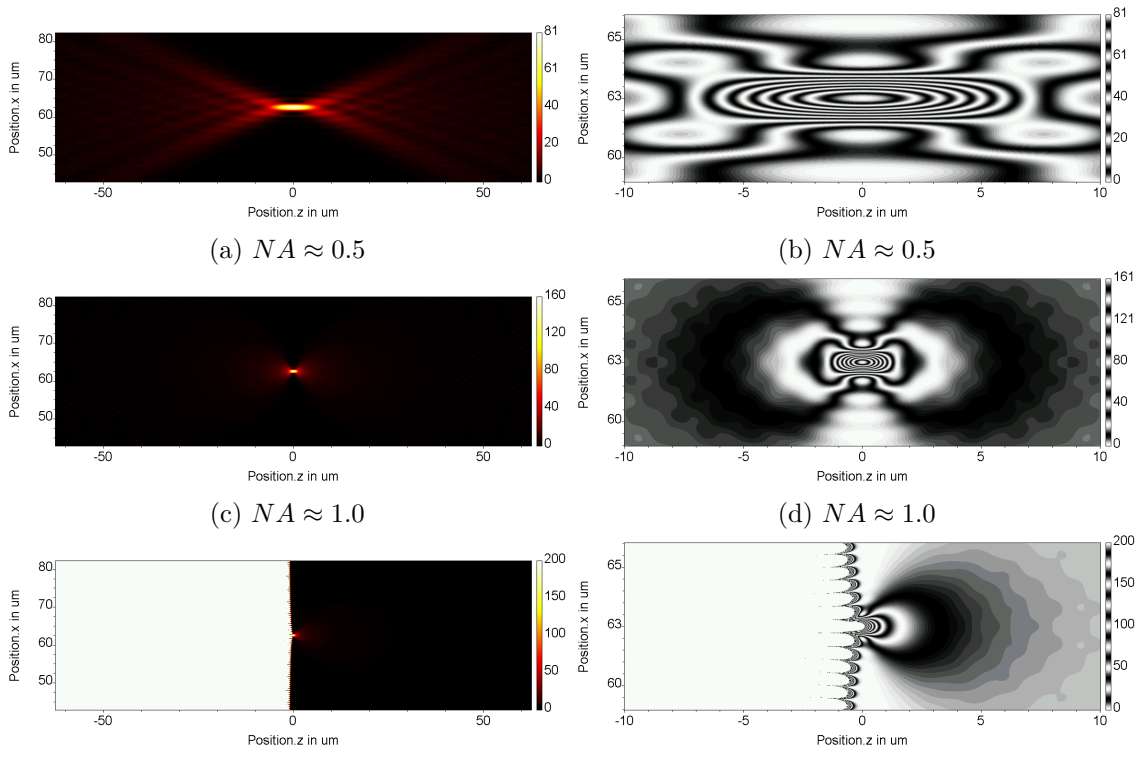


Figure 3.16: Transmission efficiency as a function of the spot position in x

3.10.3 Ideal focusing

In Sec. 3.9.1 the ASD was described as a way to express an input field as a superposition of multiple plane waves. The spectrum of these plane waves is also called k-space or directional space, in which every discrete plane wave is represented by a distinct point $(k_{i;m}, k_{i;n}) = (2\pi\ m/P_x, 2\pi\ n/P_y)$ with $k_{i;m}$ and $k_{i;n}$ related to the incident angles by Eq. 2.7. While usually it is the ASD that determines the scalar amplitude of every point in k-space in order to model an arbitrary input field in the LIF-approach, another interesting case is the full and equal occupation of k-space. In this scenario every mode or plane wave of the input field is equally excited, forming an ideal focus in superposition.
In a numerical simulation the size of the focus is limited by the truncation order. Figs. 3.17a and 3.17b show a x/z-slice of the focus with $M_0 = 40$ ($M = 2M_0 + 1$), which is half the number of propagative modes and refers to a numerical aperture of $NA \approx 0.5$. The evanescent limit is reached at $m_{\text{prop-max}} = \pm\lfloor P_x/\lambda \rfloor = \pm 80$, as in this example the following parameters were chosen: $P_x = 125\ \mu m, n_I = 1, \lambda = 1\ \mu m$. Fig. 3.17c and 3.17d show the focus spot with all propagative modes ($M_0 = 80 \Rightarrow NA \approx 1$). Allowing also evanescent waves for the construction of the focus, the size of the focus can be further decreased. However, this comes with some physical problems as illustrated in Fig. 3.17e and 3.17f, where the truncation order is $M_0 = 120$ ($\Rightarrow NA \approx 1.5$). Due to Eq. 3.2 ($\Rightarrow \mathbf{E}(z) \sim exp(i\gamma_I z)$) the



(a) $NA \approx 0.5$ (b) $NA \approx 0.5$

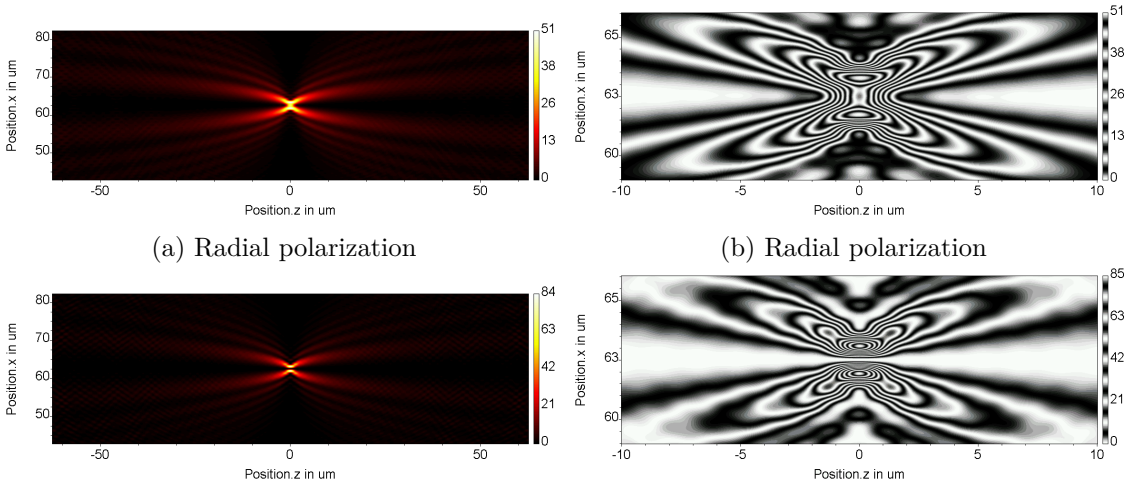
(c) $NA \approx 1.0$ (d) $NA \approx 1.0$

(e) $NA \approx 1.5$ (f) $NA \approx 1.5$

Figure 3.17: LIF-RCWA simulation of linear polarized focus spots with different numerical apertures and two color palettes for different zoom factors.

evanescent waves (with $\mathrm{Im}\,(\gamma_I) > 0$) will increase exponentially in negative z-distance from the focus leading to an unphysical light source model. Thus, for most simulations those evanescent modes in the input field should be neglected in order to use the benefits of high mode truncation in the analysis and at the same time to avoid the need for an unphysical light source.

Independently of the scalar amplitudes in the plane wave spectrum, the LIF-RCWA also allows various forms of polarization (cf. Sec. 3.9.2). While Figs. 3.17a-3.17f showed different cases of linear polarized light, focus spots with radially and azimuthally polarized light are shown in Figs. 3.18a and 3.18c.



(a) Radial polarization

(b) Radial polarization

(c) Azimuthal polarization

(d) Azimuthal polarization

Figure 3.18: LIF-RCWA simulation of radial and azimuthal polarized focus spots with full numerical apertures and two color palettes for different zoom factors.

Perfect focusing can also be used to create a virtual point light source. In Fig. 3.19 a homogeneous spacing layer with refractive index of $n_{II,0} = 1$ and thickness $d_0 = 15\,\mu m$ is added in front of the Ronchi grating that was discussed in Sec. 3.10.1. The result is a multitude of diffraction orders that appear in the far field due to the curved wavefront of the virtual point light source.

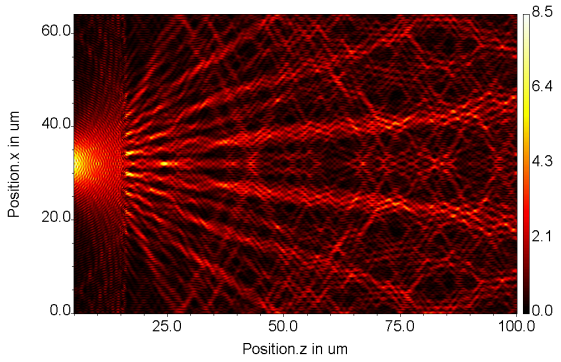


Figure 3.19: Intensity distribution of a Ronchi grating under point light illumination.

3.10.4 Waveguide coupling and propagation as another application of the LIF-RCWA

An analytical description of the propagation of light in a waveguide is not a trivial task, since it requires a new model for every new geometry. Here, the LIF-RCWA offers a powerful numerical alternative that can handle arbitrary waveguides and even describe the coupling process as well as potential coupling and leakage losses.

Planar dielectric waveguide

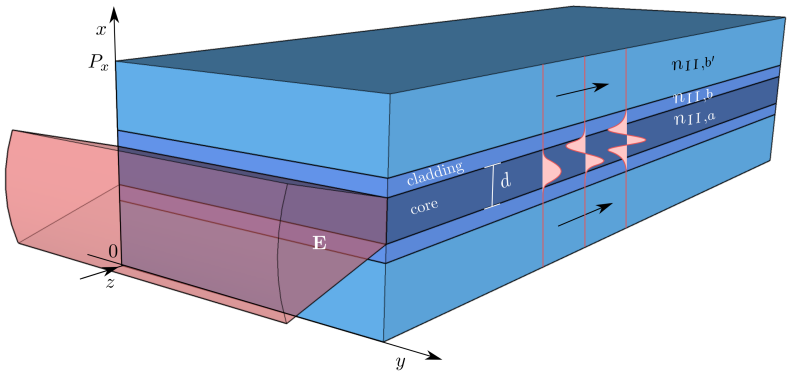


Figure 3.20: Illustration of the coupling of a Gaussian beam into a planar single-mode waveguide and propagation of a waveguide mode

In order to demonstrate the potential of the LIF-RCWA for waveguide applications, a symmetric planar dielectric waveguide as depicted in Fig. 3.20 serves as an example. It consists of a dielectric core with $n_{II,a} = 1.47$ and diameter d and a cladding material on top and below with a lower refractive index of $n_{II,b} = 1.46$.

The following analytical model is described in detail by Saleh and Teich[110]. With Snell's law and simple ray optics a critical angle θ_c can be derived below which light is guided through the core material via total internal reflection, defining a numerical aperture for the waveguide as

$$\mathrm{NA} = \sqrt{n_{II,a}^2 - n_{II,b}^2}. \tag{3.39}$$

However, only discrete incident angles are actually propagative, a fact that can only be explained when considering the wave nature of light. Due to the interference of the waves inside the waveguide, only those waves remain which are in phase with their internally reflected counterparts after one round trip - since only those happen to interfere constructively.

For TE-polarized light, this leads to the following self-consistency condition:

$$\tan\left(\pi\frac{d}{\lambda}n_{II,a}\sin\left(\theta\right)-m\frac{\pi}{2}\right)=\sqrt{\frac{\cos^2\left(\theta_c\right)}{\sin^2\left(\theta\right)}-1}, \tag{3.40}$$

which has only N discrete solutions for the incident angle θ with

$$N=\left\lfloor\frac{2d}{\lambda}\mathrm{NA}\right\rfloor. \tag{3.41}$$

For a core diameter of $d=12\,\mu m$ the analytical model thus predicts $N=3$ distinct so-called TE-modes at a wavelength of $\lambda=1.55\mu m$, each which is composed of two plane waves traveling through the waveguide at the angle θ_m and $-\theta_m$. In superposition they form the electric field

$$E_y\left(x,z\right)=\sum_m\sqrt{\eta_m}u_m\left(x\right)\exp\left(ik_{z;m}z\right), \tag{3.42}$$

where η_m is a constant amplitude and u_m describes the field distribution of the individual modes in lateral direction. Inside the waveguide's core ($-d/2<x<d/2$) the field turns out to be

$$u_m\left(x\right)=\begin{cases}\cos\left(k_{x;m}x\right)=\cos\left(\dfrac{2\pi}{\lambda}\sin\left(\theta_m\right)x\right), & m=0,2,4,...\\ \sin\left(k_{x;m}x\right)=\sin\left(\dfrac{2\pi}{\lambda}\sin\left(\theta_m\right)x\right), & m=1,3,5,...\end{cases}, \tag{3.43}$$

whereas it declines exponentially in the cladding:

$$u_m\left(x\right)=\begin{cases}\exp\left(-\gamma_m x\right), & x>d/2\\ \exp\left(\gamma_m x\right), & x<-d/2\end{cases} \tag{3.44}$$

All three modes ($m = 0, 1, 2$) of this setup at $\pm\theta_0 = \pm 2.024\,\text{deg}$, $\pm\theta_1 = \pm 4.006\,\text{deg}$ and $\pm\theta_2 = \pm 5.852\,\text{deg}$ are depicted in Fig. 3.21.

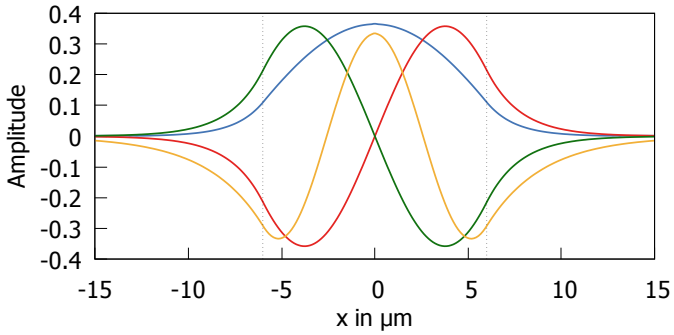
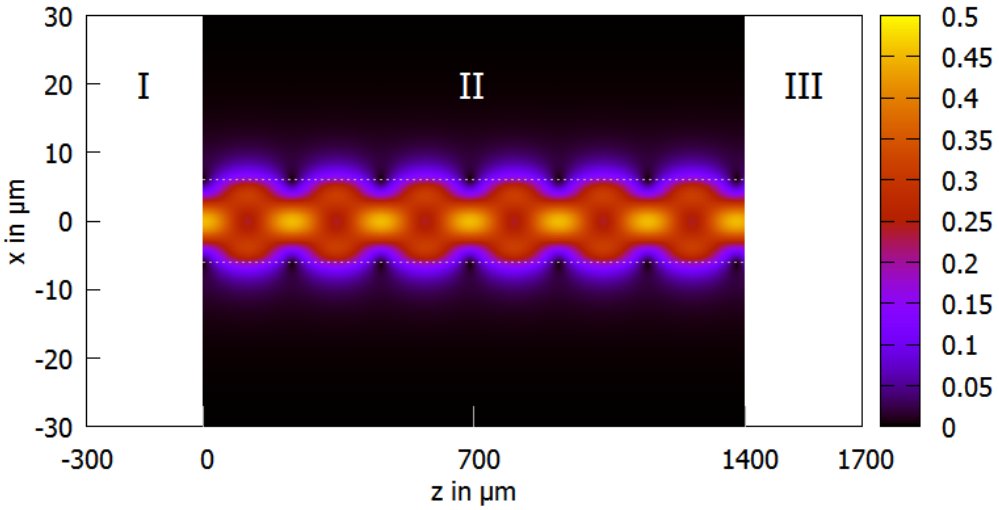


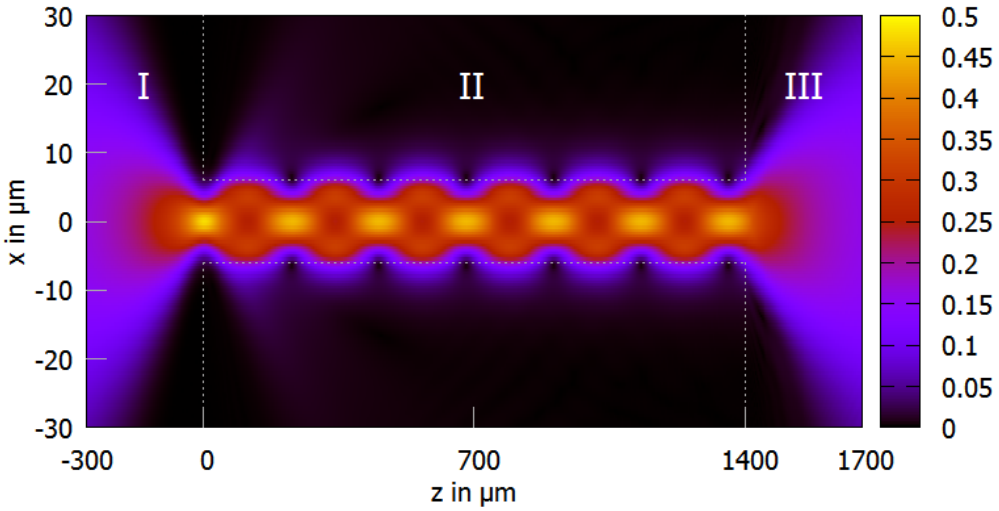
Figure 3.21: First fundamental modes of a planar dielectric waveguide: $m = 0$ in blue, $m = \pm 1$ in green and red and $m = 2$ in yellow

Their propagation along z reveals a distinct pattern in the field distribution, which can be observed in Fig. 3.22a.

The model for the numerical approach includes an incident region as well as a transmission region, in which a linearly TE-polarized Gaussian beam (cf. 3.9.1) with a waist size of $\sigma_0 = 3.5\mu m$ is coupled into the waveguide and transmits at the end. In order to avoid internal reflection on both sides of the waveguide and additional interference as a consequence, the refractive indices of the surroundings are matched to the core: $n_I = n_{III} = n_{II,a} = 1.47$. In order to avoid crosstalk between different periods - since in the RCWA every diffraction problem ultimately repeats itself in lateral direction - the period is chosen sufficiently big with $P_x = 125\,\mu m$ and complemented with an additional weak absorbing boundaries at a distance of 25 μm around the center of the core ($\Rightarrow d_{\text{cladding}} = 50\ \mu m, n'_{II,b} = 1.46 + 8.56 \cdot 10^{-4} i$, cf. Fig. 3.20).
The numerical calculation was performed with the LIF-RCWA and the result can be observed in Fig. 3.22b. Between $z = 0\,\mu m$ and $z = 1400\,\mu m$ it shows the electric field inside the waveguide matching the pattern of the analytical model. In addition it shows the coupling of the Gaussian beam into the waveguide at $z < 0\,\mu m$ as well as the transmission into region *III*.



(a) Field distribution according to an anaytical model



(b) Field distribution according to the LIF-RCWA

Figure 3.22: Propagation of the three fundamental modes of a planar dielectric multi-mode waveguide with a core diameter 12 μm, calculated with an analytical model (a) and the LIF-RCWA (b)

Finally, also the coupling losses due to a mismatch between the Gaussian incident beam and the propagative waveguide modes are visualized. The degree of matching between the two can be determined by an overlap integral

$$\eta\left(\sigma_0\right) = \frac{\left|\int_{-P_x/2}^{P_x/2} E_y^{\mathrm{GB}}\left(x,\sigma_0\right) E_{y,m}^{\mathrm{WM}}\left(x\right)^* \mathrm{dx}\right|^2}{\int_{-P_x/2}^{P_x/2}\left|E_y^{\mathrm{GB}}\left(x,\sigma_0\right)\right|^2 \mathrm{dx} \int_{-P_x/2}^{P_x/2}\left|E_{y,m}^{\mathrm{WM}}\left(x\right)\right|^2 \mathrm{dx}}, \tag{3.45}$$

where E_y^{GB} describes the profile of the Gaussian beam at the plane in the origin, whereas $E_{y,m}^{\mathrm{WM}}$ denotes the m-th propagative waveguide mode.

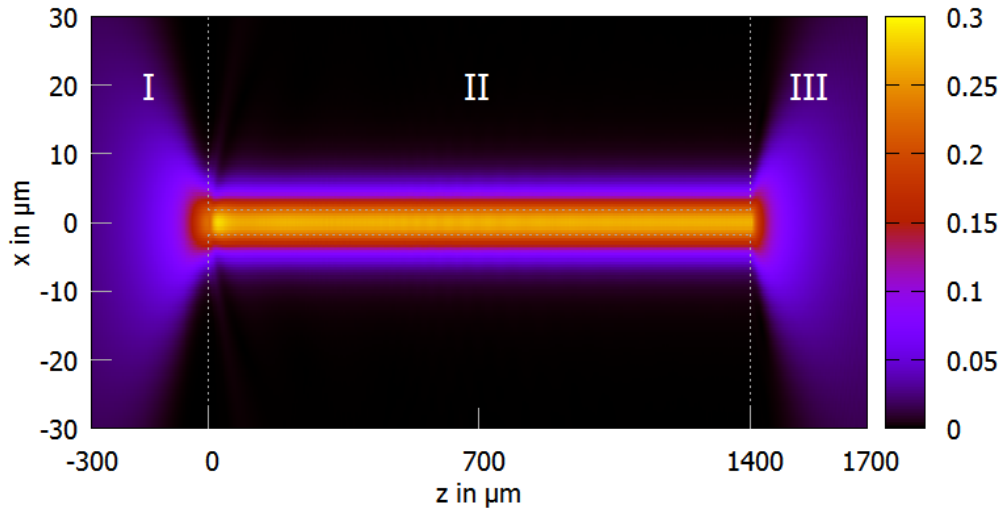


Figure 3.23: Field distribution in a planar dielectric single-mode waveguide with core diameter $3.72\,\mu m$, calculated by the LIF-RCWA

In the numerical calculation only the propagative modes account for the transmitted light in region *III* since all leaky modes should be absorbed by the additionally inserted boundaries. In a single-mode setup with core diameter $d = 3.72\,\mu m$ as depicted in Fig. 3.23 and with no reflections in play, the overlap integral of the incident Gaussian beam and the only analytically derived propagative waveguide mode should be comparable to the transmission efficiency of th LIF-RCWA. And in fact they match perfectly for various values of the Gaussian beam waist size σ_0 (cf. Eq. 3.30). These results have been published in [4].

In conclusion, both, the coupling as well as the formation and propagation of distinct waveguide modes, have been demonstrated to be in good agreement with standard theories for a single- and a multi-mode waveguide. This demonstrates the validity of the LIF-RCWA as a method to rigorously analyze optical waveguides.

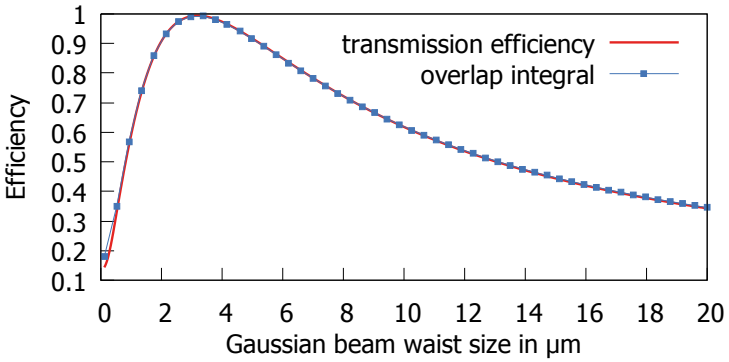


Figure 3.24: Comparison between the waveguide's transmission efficiency (solid blue squares) calculated by the LIF-RCWA and the overlap integral (solid red line) of a Gaussian beam profile with the waveguides single-mode

Multi-mode interference optical splitter

Macroscopic waveguides can be utilized to work as interference optical splitter. Fig. 3.25 shows (in anisotropic resolution) the coupling of a Gaussian beam with a waist size of $\sigma_0 = 10\,\mu m$ and a wavelength of $\lambda = 0.85\,\mu m$ under an angle of $\theta = 5$ deg in such a macroscopic multi-mode waveguide. The refractive indices are equal to the single-mode waveguide in the previous example, but this time the core has a diameter of $d_{\text{core}} = 100\,\mu m$ and the simulation period is $P_x = 1\,mm$. After several occasions of internal reflections at the waveguide cladding a fan out into multiple discrete modes can be observed. In a real application those outputs can be further routed to distribute a broadcast signal.

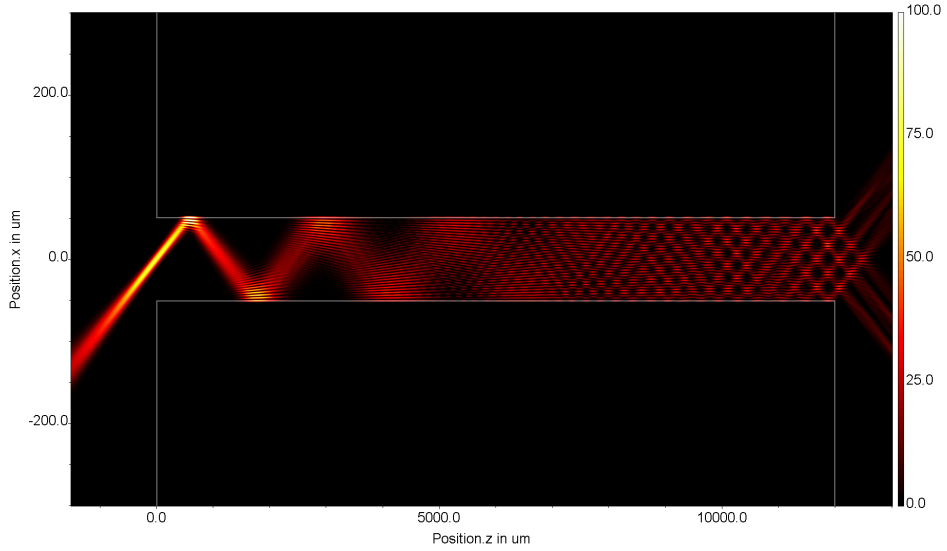


Figure 3.25: Intensity distribution of a linear polarized Gaussian beam with a beam waist of $\sigma_0 = 10\,\mu m$ and wavelength of $\lambda = 0.85\,\mu m$ coupling into a multi-mode interference optical splitter with a core width of $d_{\text{core}} = 100\,\mu m$ under an incidence angle of $\theta = 5$ deg (anisotropic resolution), $M = 402$ modes

3.11 Conclusion

In this chapter a novel approach has been presented that enables the rigorous coupled-wave analysis (RCWA) to study diffraction problems with localized input fields, while maintaining its computational complexity. The detailed description covers the modeling of coherent light sources, like the Gaussian beam, an ideal focus as well as a point light source and finally the angular spectrum decomposition (ASD) as a recipe to convert an arbitrary electromagnetic field into a suitable form for the localized input field approach (LIF-RCWA) that was presented here. In order to unfold the full potential of the rigorous vector theory, different types of polarization like azimuthal and radial polarization have also been discussed, which are not available in the standard RCWA.
In various examples, the potential of the method and new possible fields of applications for the RCWA like the rigorous analysis of waveguide coupling or the simulation of the read-out process of optical medias have been demonstrated. At the same time the correctness of the new method has been verified against the standard RCWA algorithm and other theories.

Furthermore, a close look at the truncation scheme, which is used in the standard RCWA revealed an inconsistent bandwidth limitation that has been overcome in the presented method. In a comparison of convergence speeds for a non-bandlimited grating in TE-polarization, the new approach was shown to fall behind the standard RCWA. In TM-polarization, however, the new method showed superior performance - without the need and use of the inverse rule[2] - as it manages to avoid the numerical instabilities caused by the Gibbs phenomenon. Both results are founded in the consistent use of a discrete Fourier transformation with an equal number of sampling points in both space and frequency domain. However, it should be mentioned, that at the cost of the newly gained consistency in the bandlimiting, the inverse rule can also be effectively applied to the LIF-RCWA when using a different number of sampling points in real space and Fourier space. In this case the convergence speed of the LIF-RCWA, with its ability to treat non-plane wave incident light, exactly matches the standard approach for all states of polarization.
Although the simulations presented in this chapter only considered two-dimensional grating structures, this is by no means a limitation of the method. In fact, all its considerations address the analysis of the full three-dimensional diffraction problem, even though some discussions sticked to the two-dimensional case to avoid unnecessary complexity. The only reason why tree-dimensional problems have been excluded from the examples, lies in the sheer magnitude of the computational effort of those calculations. And thus the simulations could not be performed with a sufficient order of truncation on a standard desktop computer to reveal valuable insights.

[2] Avoiding the need for the inverse rule especially reduces complexity when studying two-dimensional grating structures, where the formulation becomes somewhat cumbersome (cf. Sec. 2.5.4).

Chapter 4

Near field calculation

Typically, gratings are optimized for high diffraction efficiencies in the far field as they serve as wavelength separator devices in monochromators, spectrometers and in the field of optical data processing. However, in some applications near-field patterns are the relevant quantities. Their visualization can give valuable insights in the underlying physics, e.g. when optimizing resonant structures. Electromagnetic near-fields also directly influence local absorption, which is discussed and utilized in Chap. 5 and 6. Therefore an exact calculation of the near-fields is inevitable.

In the current chapter, different methods for the calculation of near-fields in one- and two-dimensional gratings are compared. Since the diffraction coefficients of all RCWA variants presented so far were shown to have congruent convergent limits, it can be assumed, that for a sufficiently high number of Fourier modes, the electromagnetic fields are unambiguous. For smaller mode counts, however, the question remains how the electromagnetic fields of the different RCWA variants differ and how well they withstand physical consistency validation. The answer to this question is indeed important, since, due to its Fourier nature, the RCWA has intrinsic difficulties modeling lateral jump discontinuities in the material properties as they leads to high frequencies and require high truncation limits. Furthermore, runtime and memory constraints often prevent the analysis of fully converged fields. This is especially true for the three-dimensional RCWA as shown in Sec. 3.7.

4.1 Mathematical derivation

The first part of this chapter concentrates on the derivation of an efficient calculation method for the electromagnetic fields as they follow from the diffraction coefficients that are provided by the RCWA.
All following considerations apply for volume gratings with locally varying permeability and make use of the inverse rule (cf. Sec. 2.5.2). Thus, simpler forms of the method, like for one-dimensional gratings, neutral relative permeability ($\mu = 1$) or without the use of the inverse rule can easily be derived.

4.1.1 Electric and magnetic field in the outside regions

According to the Eqs. 2.6a to 2.6d, the electric field in the incident region I is defined as

$$E_{I;j}(\mathbf{r}) = E_{1;j} e^{i(k_{i;x}x + k_{i;y}y + k_{i;z}z)} + \sum_{m,n} R_{j;m,n} e^{i\left(k_{x;m}x + k_{y;n}y - \gamma_{I;m,n}z\right)} \tag{4.1}$$

for every field component $j = x, y, z$. Using the following abbreviations
$O_0(x,y) := \exp(i(k_{i;x}x + k_{i;y}y))$ and $P^+_{I,0}(z) := \exp(ik_{i;z}z)$ as well as
$\mathbf{O}(x,y) := \operatorname{diag}(\exp(i(k_{x;m}x + k_{y;n}y)))$ and $\mathbf{P}^-_I(z) := \exp(-i\gamma_{I;m,n}z)$ Eq. 4.1 reads as

$$E_{I;j}(\mathbf{r}) = O_0(x,y)\, P^+_{I,0}(z)\, e_{E_i;j} + \mathbf{O}(x,y) \operatorname{diag}\left(\mathbf{P}^-_I(z)\right) \mathbf{R}_j. \tag{4.2}$$

Due to Snell's law, $k_{x;m}$ and $k_{y;n}$ remain unchanged in the entire simulation area. Thus, for a given set of lateral coordinates $x_q = x_0..x_{N_x-1}$ and $y_r = y_0..y_{N_y-1}$ the Fourier matrix

$$\mathbf{O}^{[x,y]}_{r\cdot N_x+q;n\cdot M_x+m} = e^{i(k_{x;m}x_q + k_{y;n}y_r)} \tag{4.3}$$

must only be calculated once and can then be reused in order to save calculation time. (Here, $M_x \times M_y$ is the number of sampling points in Fourier space and $N_x \times N_y$ is the number of sampling points in position space respectively.) This way the electric field in a slice of discrete points in space at position z can be calculated at once by

$$\mathbf{E}^{[x,y]}_{I;j}(z) = \mathbf{O}^{[x,y]} \left(\operatorname{diag}\left(\mathbf{P}^+_I(z)\right) \mathbf{L}_{0;j} + \operatorname{diag}\left(\mathbf{P}^-_I(z)\right) \mathbf{R}_j\right), \tag{4.4}$$

with the incident vector $\mathbf{L}_0$ as defined according to definition 2.26b.

In the LIF-approach the corresponding electric field in region I takes a very similar but even more symmetric form:

$$\mathbf{E}^{[x,y]}_{I;j}(z) = \mathbf{O}^{[x,y]} \left(\operatorname{diag}\left(\mathbf{P}^+_I(z)\right) \mathbf{L}_j + \operatorname{diag}\left(\mathbf{P}^-_I(z)\right) \mathbf{R}_j\right) \tag{4.5}$$

Remark: It should be noted here, that in a computer algorithm, it is of course sufficient to store a vector instead of a diagonal matrix with a square number of values. As a consequence,

also the matrix-vector product of a diagonal matrix and a vector can be implemented very efficiently.

While the tangential diffraction coefficients are directly provided by the RCWA-algorithm, the z-components in $\mathbf{R}_z$ derive from the orthogonality of every diffracted mode and its wave vector $\mathbf{k}_{I;m,n}\mathbf{R}_{m,n} = 0$ leading to

$$\mathbf{R}_z = \mathbf{K}_{I,z}^{-1} \left(\mathbf{K}_x \mathbf{R}_x + \mathbf{K}_y \mathbf{R}_y \right). \text{[1]} \tag{4.6}$$

The very same considerations also apply to the electric field in the transmitting region (*III*):

$$\mathbf{E}_{III;j}^{[x,y]}(z) = \mathbf{O}^{[x,y]} \operatorname{diag}\left(\mathbf{P}_{III}^{+}(z) \right) \mathbf{T}_j \tag{4.7}$$

with

$$\mathbf{T}_z = \mathbf{K}_{III,z}^{-1} \left(\mathbf{K}_x \mathbf{T}_x + \mathbf{K}_y \mathbf{T}_y \right). \tag{4.8}$$

And after transforming the tangential diffraction coefficients of the electric field into its magnetic counterparts by using the conversion matrix $\mathbf{C}$ (cf. 2.28):

$$\mathbf{L}_{\perp}^{H} = \mathbf{C}\mathbf{L}_{\perp} \tag{4.9a}$$
$$\mathbf{R}_{\perp}^{H} = \mathbf{C}\mathbf{R}_{\perp} \tag{4.9b}$$
$$\mathbf{T}_{\perp}^{H} = \mathbf{C}\mathbf{T}_{\perp} \tag{4.9c}$$

Eqs. 4.4 to 4.8 also apply equally to the magnetic fields.

[1] As introduced in Chap. 2, $\mathbf{K}_j$ is the diagonal matrix of $k_{j;m,n}$.

4.1.2 Electric and magnetic field in the grating region

In any layer l in the grating region (II) the electric and the magnetic field are defined by the partial Fourier series according to Eqs. 2.9a and 2.9b:

$$E_{l;j}(\mathbf{r}) = \sum_{m,n} S_{l;j;m,n}(z) e^{i(k_{x;m}x + k_{y;n}y)} \tag{4.10a}$$

$$H_{l;j}(\mathbf{r}) = \frac{1}{Z_0} \sum_{m,n} U_{l;j;m,n}(z) e^{i(k_{x;m}x + k_{y;n}y)} \tag{4.10b}$$

Using again the O-matrix notation of Eq. 4.3 leads to

$$\mathbf{E}_{l;j}^{[x,y]}(z) = \mathbf{O}^{[x,y]} \mathbf{S}_{l;j}(z) \tag{4.11a}$$

$$\mathbf{H}_{l;j}^{[x,y]}(z) = \frac{1}{Z_0} \mathbf{O}^{[x,y]} \mathbf{U}_{l;j}(z) \tag{4.11b}$$

with the $\mathbf{S}_{l;\perp}$ and $\mathbf{U}_{l;\perp}$ defined according to Eqs. 2.21a and 2.21b:

$$\mathbf{S}_{l;\perp}(z) = \mathbf{V}_l \left(\mathbf{P}_l^+(z)\ \mathbf{t}_l + \mathbf{P}_l^-(z)\ \mathbf{r}_l \right) \tag{4.12a}$$

$$\mathbf{U}_{l;\perp}(z) = \mathbf{W}_l \left(\mathbf{P}_l^+(z)\ \mathbf{t}_l - \mathbf{P}_l^-(z)\ \mathbf{r}_l \right). \tag{4.12b}$$

In Chap. 2 Maxwell's equations were solved by forward and backward propagating eigenmodes for the grating region rather than by Fourier modes of the electric and magnetic field. As a consequence, the z-components of the wave vectors, one might extract from $\mathbf{P}_l^+(z) = \exp\left(-k_0\sqrt{\lambda_q}z\right) \Rightarrow k_{z;q} = ik_0\sqrt{\lambda_q}$), are not compatible with the wave vector's tangential components of the Fourier modes, and so they cannot be used to derive the z-components of the fields in region II as in Eq. 4.6 and 4.8. However, using Maxwell's equations, the z-component of the electric field can be expressed by the tangential components of the magnetic field and vice versa:

$$E_z = \frac{i}{\omega\epsilon\epsilon_0} \left(\frac{\partial}{\partial x} H_y - \frac{\partial}{\partial y} H_x \right) \tag{4.13a}$$

$$H_z = \frac{-i}{\omega\mu\mu_0} \left(\frac{\partial}{\partial x} E_y - \frac{\partial}{\partial y} E_x \right) \tag{4.13b}$$

Substituting the Fourier ansatz for the permittivity and the fields (cf. Eqs. 2.1, 2.9a and 2.9b) and again using the **O** matrix notation leads to

$$\mathbf{O}^{[x,y]}\mathbf{S}_{l;z}(z) = \frac{iZ_0}{k_0}\,\mathbf{O}^{[x,y]}\left(\lceil\lfloor\epsilon_l\rfloor\rceil\mathbf{K}_x\frac{1}{Z_0}\mathbf{U}_{l;y}(z) - \lfloor\lceil\epsilon_l\rceil\rfloor\mathbf{K}_y\frac{1}{Z_0}\mathbf{U}_{l;x}(z)\right) \tag{4.14a}$$

$$\frac{1}{Z_0}\mathbf{O}^{[x,y]}\mathbf{U}_{l;z}(z) = \frac{i}{k_0 Z_0}\mathbf{O}^{[x,y]}\left(\lceil\lfloor\mu_l\rfloor\rceil\mathbf{K}_x\mathbf{S}_{l;y}(z) - \lfloor\lceil\mu_l\rceil\rfloor\mathbf{K}_y\mathbf{S}_{l;x}(z)\right) \tag{4.14b}$$

and finally to an expression for the z-components of the Fourier ansatz in the grating region (*II*):

$$\mathbf{S}_{l;z}(z) = \frac{i}{k_0}\begin{pmatrix}0 & \lceil\lfloor\epsilon_l\rfloor\rceil\\ -\lfloor\lceil\epsilon_l\rceil\rfloor & 0\end{pmatrix}\begin{pmatrix}\mathbf{K}_y\\ \mathbf{K}_x\end{pmatrix}\mathbf{U}_{l;\perp}(z) \tag{4.15a}$$

$$\mathbf{U}_{l;z}(z) = \frac{i}{k_0}\begin{pmatrix}0 & \lceil\lfloor\mu_l\rfloor\rceil\\ -\lfloor\lceil\mu_l\rceil\rfloor & 0\end{pmatrix}\begin{pmatrix}\mathbf{K}_y\\ \mathbf{K}_x\end{pmatrix}\mathbf{S}_{l;\perp}(z) \tag{4.15b}$$

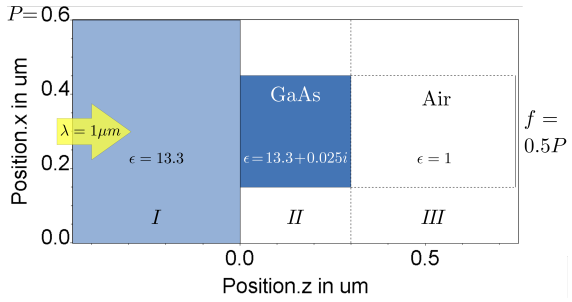
Now, every component of the electric and the magnetic near field can be calculated in any arbitrary simulation area.

Example and discussion

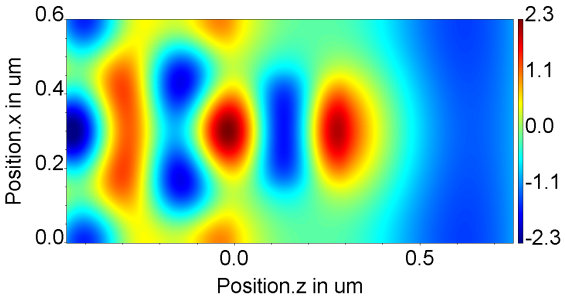
Fig. 4.1a shows the sketch of a GaAs-Ronchi grating with a duty cycle of $f = 0.5\,P$ and a period of $P = 0.5\,\mu m$. The permittivity of GaAs is assumed to be $\epsilon = 13.3 + 0.025i$ and the illumination light has a wavelength of $\lambda = 1\,\mu m$. In Figs. 4.1b to 4.1d the different non-zero components of the electric near field are calculated for TE and TM polarization. A high mode count ($M = 301$) was chosen in order to obtain a reference solution for later comparisons.

4.2 Validation of the electromagnetic fields

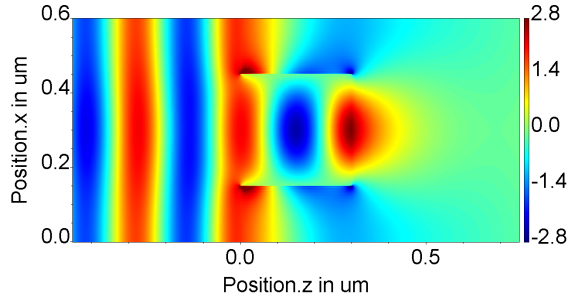
It assumed that for a sufficiently high mode count as it is shown in Fig. 4.1 the electromagnetic fields of all discussed RCWA variants are converged equally to the same correct physical near fields. It is also obvious that the electromagnetic fields of the RCWA variants will differ for smaller mode counts. However, especially complex diffraction problems in the three-dimensional regime, often require a very high number of modes to achieve reasonable results and may exceed the available computational resources. This leads to the question if the electromagnetic fields converge with the same speed as their diffraction coefficients. They can clearly not converge faster but possibly slower, if the error in the diffraction coefficients multiplies in the calculation of the fields. Another question is how to validate the correctness of the fields.



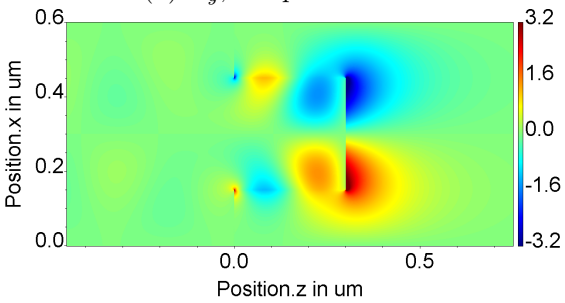
(a) Sketch of a GaAs-Grating



(b) E_y, TE polarization



(c) E_x, TM polarization



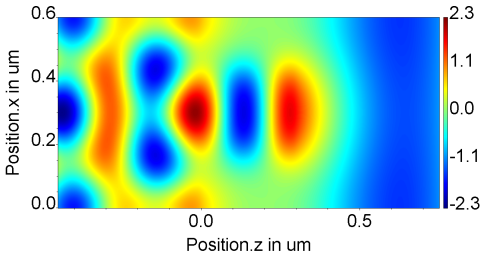
(d) E_z, TM polarization

Figure 4.1: Near field calculation with $M_0 = 150 \Rightarrow M = 301$ Fourier modes

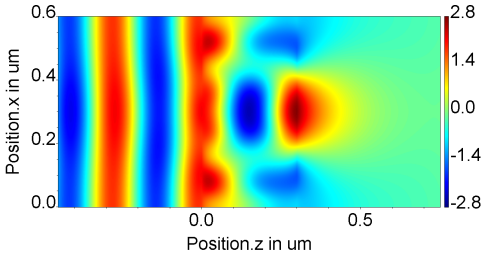
To begin with, Fig. 4.2 shows the electric near fields as they are calculated according to the derivation in the previous section but on the base of the diffraction coefficients of (1) the standard RCWA, (2) the RCWA with inverse rule applied (cf. Li's rules in Sec. 2.5.2), (3) the RCWA with apodization and (4) with the LIF-RCWA, all at a very low mode count of only $M_0 = 4 \Rightarrow M = 9$. Only Fig. 4.2a shows the y-component of the electric field in TE polarization, since it is obviously already converged at this low mode count and is therefore not the focus of the following discussion.

As expected, the fields of all RCWA variants slightly differ, but they all show the difficulty in the Fourier expanded x-component of the electric field to describe steep slopes at low mode counts. The field calculated with an additional apodization (cf. Figs. 4.2f) appears even smoother due to the suppressed higher frequencies. Since $N = 512$ sampling points were used in all these simulations for the discrete Fourier transformation of the permittivity, also the electric fields are reconstructed on a discrete grid with 512 pixels per grating period. By contrast, the LIF-RCWA, uses the same number of sampling points in Fourier and position space ($\Rightarrow N = M = 9$). But, in order to obtain comparable results, in this example the fields of the LIF-RCWA are equally reconstructed on $N = 512$ discrete positions in x-direction (cf. Figs. 4.2h and 4.2i).

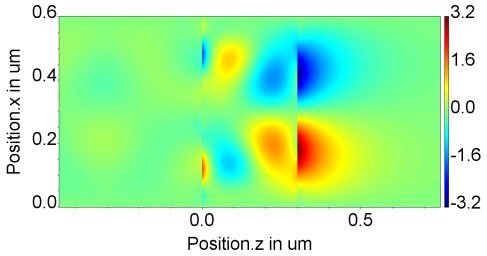
It is obviously not possible to decide, which of the methods in Fig. 4.2 provides the *best* solution for the electromagnetic near fields based on the visual impression. So there is an apparent need for some substantial criteria.



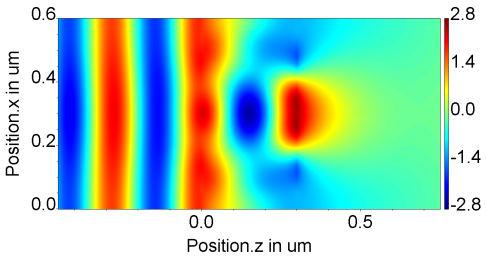
(a) E_y, TE polarization, Std-RCWA



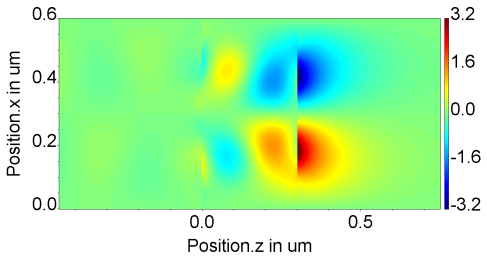
(b) E_x, TM polarization, Std-RCWA



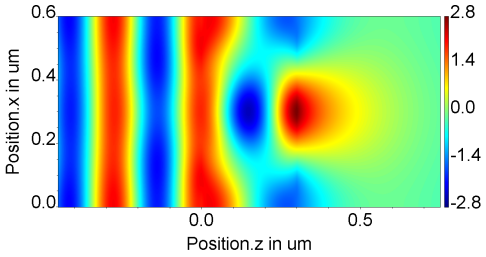
(c) E_z, TM polarization, Std-RCWA



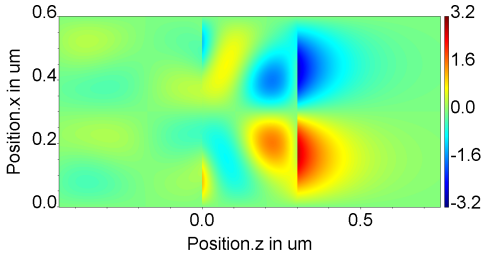
(d) E_x, TM polarization, RCWA with inverse rule applied



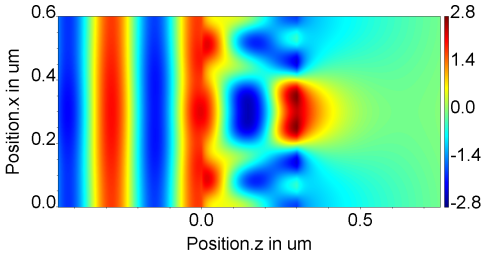
(e) E_z, TM polarization, RCWA with inverse rule applied



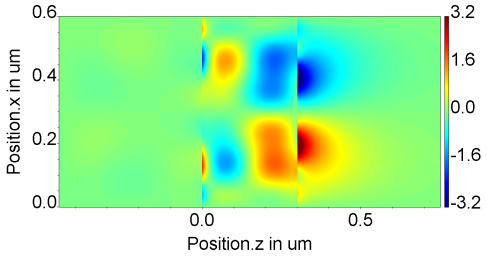
(f) E_x, TM polarization, RCWA with apodization



(g) E_z, TM polarization, RCWA with apodization



(h) E_x, TM polarization, LIF-RCWA



(i) E_z, TM polarization, LIF-RCWA

Figure 4.2: Electric near fields with different RCWA variants at low Fourier mode count of $M_0 = 4 \Rightarrow M = 9$ reconstructed on 512 pixels in x-direction.

4.2.1 Maxwell's boundary conditions

As indicated before, one important requirement that every near field solution is expected to meet is its agreement with Maxwell's boundary conditions (cf. Sec. 1.1.7). According to this, the components of the electric and the magnetic fields (E and H), which are tangential to the grating boundaries and the components of the electric displacement field and magnetic induction (D and B), which are normal to the grating boundaries, are supposed to be continuous. The remaining components, however, might be discontinuous and in case of material jump discontinuities, they are expected to be so. This follows straight from the constitutive equations (e.g. $D = \epsilon\epsilon_0 E$, cf. Eqs. 1.2a and 1.2a).
The conditions for the individual electric field components are shown in Tab. 4.1 for a one-dimensional grating with periodicity in x-direction in non-conical mounting. At TE polarization, E_y is the only non-zero component of the electric field and it is tangential to any boundary, which is encountered in x- and z-direction. The resulting continuity requirement in the same directions are indicated by the letter "c". Since Fig. 4.2a shows both continuities, the corresponding cells are marked in green. The y-component would be discontinuous at any boundary encountered in y-direction, which is indicated by the letter "d". But only a one-dimensional grating is considered here, and thus, the continuity of the field in y-direction does not violate Maxwell's conditions.
In the TM case, the conditions only require the continuity of E_x in y- and z-direction and due to the material discontinuity in x-direction E_x is in fact expected to follow these discontinuities. But because in x-direction the electric field is described by a Fourier series with a finite number of Fourier modes, the calculated field of any of the methods (cf. Fig. 4.2b to 4.2i) and especially of the methods, which use apodization and the inverse rule, is continuous. This is a violation of discontinuity condition and indicated by a red cell color.
The z-component of the electric field only allows a discontinuity in z-direction, which can be observed in all the results (cf. Fig 4.2c, 4.2e and 4.2i) on the right interface, where the GaAs-grating faces region *III* with a refractive index of $n = 1$.

TE polarization

	x	y	z
$E_x = 0$			
E_y	c	d	c
$E_z = 0$			

TM polarization

	x	y	z
E_x	d	c	c
$E_y = 0$			
E_z	c	c	d

Table 4.1: Maxwell's boundary conditions

There are different solutions to this problem: Since all Fourier methods lack the ability to model steep slopes at finite mode counts, one could think of using different base functions. In fact, Khavasi et al.[62], for example, rewrote the RCWA using a Legendre polynomial expansion instead of the common Fourier expansions - also with the goal to avoid the Gibbs phenomenon in the TM case and its negative effects the numerical stability. An approach based on Fourier expansion was proposed by Lalanne and Jurek[67], who introduced a non-bandwidth limited quantity to the calculation of the fields. Finally, apodization (cf. Sec. 2.6) aims at avoiding the problem completely by changing the design properties of the grating to obtain a structure with smoother material transitions and a smaller bandwidth. Thus,

the critical boundary condition is met not by making the x-component of the electric field discontinuous in x-direction, but by changing the conditions.

The method of Lalanne and Jurek

TM polarization

	x	y	z
D_x	**c**	d	d
$D_y = 0$			
D_z	d	d	c

	x	y	z
E_x	**d**	d	d
$E_y = 0$			
E_z	c	c	d

Table 4.2: Maxwell's boundary conditions

Lalanne and Jurek[67] faced the discontinuity violation by only calculating the x-continuous fields with a Fourier expansion and deriving all x-discontinuous fields by multiplication with the non-bandwidth limited permittivity in position space. More precisely, they suggested to calculate the x-continuous quantities E_y and E_z as usual by the Fourier series

$$E_j = \sum_m S_{j;m}(z)\, e^{ik_{x;m}x} \quad \text{for} \quad j = y, z, \tag{4.16}$$

but obtain the discontinuous x-component of the E-field from the continuous x-component of the D-field (cf. Tab. 4.2) and the discontinuous permittivity $\epsilon(x)$:

$$E_x = 1/\epsilon(x) \cdot D_x \tag{4.17}$$

with

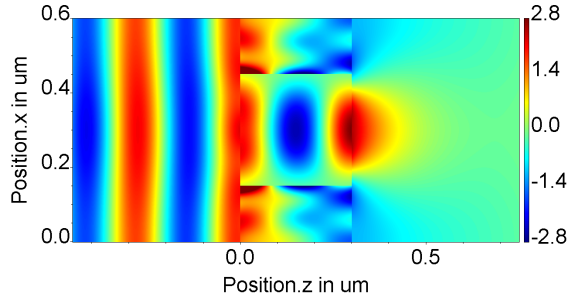
$$D_x = \sum_m G_{x;m}(z)\, e^{ik_{x;m}x}. \tag{4.18}$$

This is possible, since in the two-dimensional RCWA the D-field is directly accessible via the magnetic field: When solving $\delta^2 \mathbf{U}_y/\delta z^2 = \mathbf{\Omega}^H \mathbf{U}_y$ with an eigenvalue ansatz for the magnetic field[2] then it is $\mathbf{U}_y = \mathbf{W}\left(\mathbf{P}^+(z)\,\mathbf{t} + \mathbf{P}^-(z)\,\mathbf{r}\right)$ and the Ampère's circuital law (cf. Eq. 1.1a) leads to

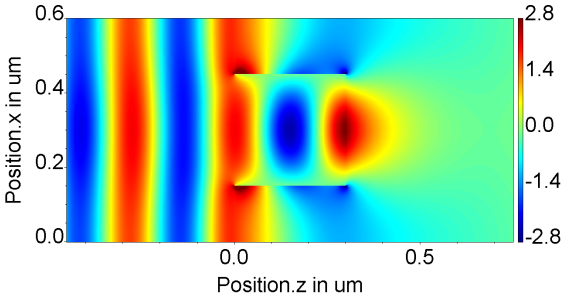
$$\mathbf{G} = \frac{ik_0}{\omega}\mathbf{W}\mathbf{Q}\left(-\mathbf{P}^+\mathbf{t} + \mathbf{P}^-\mathbf{r}\right). \tag{4.19}$$

[2]Note, that in Sec. 2.4.2 the eigenvalue ansatz was applied to the electric field (cf. Eq. 2.16). Both cases are compared in Sec. A.5, where also this two-dimensional, non-conical mounting is considered.

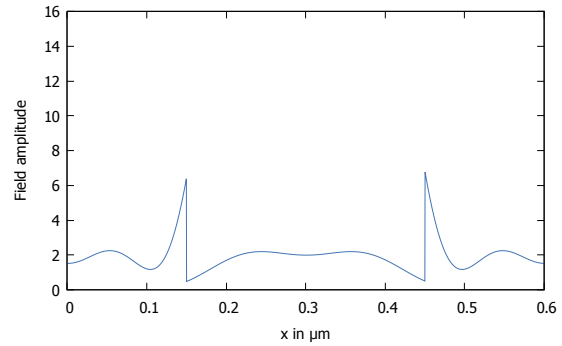
The result is shown in Fig. 4.3a and 4.3b for 4 and 150 Fourier modes. Here, the x-component of the electric field shows a sharp discontinuity in x-direction even at the small mode count, but in return suffers from strong oscillations, which disappear again at a higher mode count. More serious, however, are high amplitude peaks, which appear near the boundaries as it can be observed in Figs. 4.3c and 4.3d. These become narrower but also higher for increasing mode counts. The amplitudes in Figs. 4.3a and 4.3b are therefore clipped so they can be compared with the other results.
Another downside of this approach is that the method cannot be applied to two-dimensional structures, since in the general case the D-field is not continuous in both tangential directions (cf. Tab. 4.2). So the E-field cannot be derived.



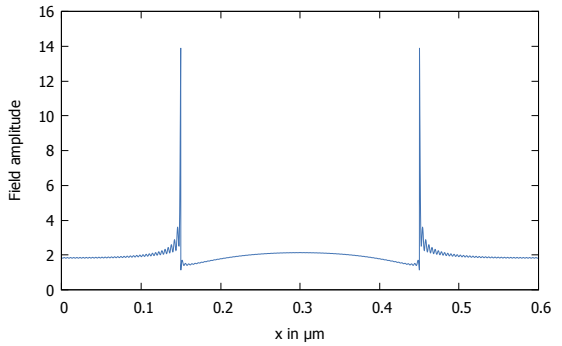
(a) E_x, TM polarization, 4 modes



(b) E_x, TM polarization, 150 modes



(c) E_x at $z = +0$, 4 modes



(d) E_x at $z = +0$, 150 modes

Figure 4.3: For a one-dimensional grating, the electric fields calculated with the method of Lalanne and Jurek[67] (with inverse rule applied) show the expected jump discontinuity in normal direction to the grating slopes but singularities at the grating boundaries.

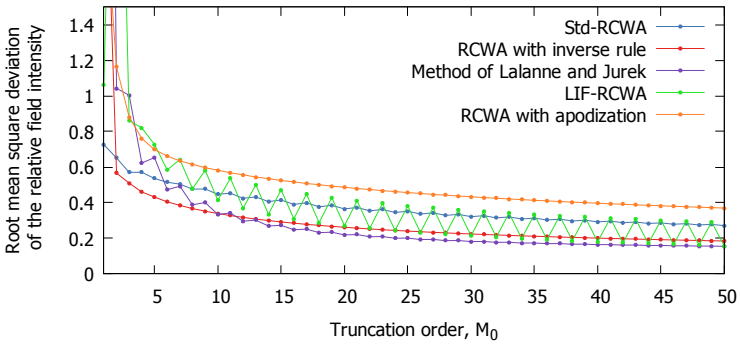
In summary, there are different methods for the calculation of the electromagnetic fields, which all lead to different solutions for small mode counts. Since the expected discontinuity of the fields components in normal direction to the grating slopes cannot be represented by a finite mode count, Lalanne and Jurek[67] suggested to achieve the discontinuity by introducing a non-bandwidth limited permittivity to the calculation. This cures the continuity problem, but the resulting fields still seem to deviate visually from the standard high-mode calculation due to singularities that arise near the surface boundaries. Thus, it turns out that Maxwell's continuity conditions are not a sufficient criterion to decide, which method is the most accurate.

4.2.2 Root mean square deviation

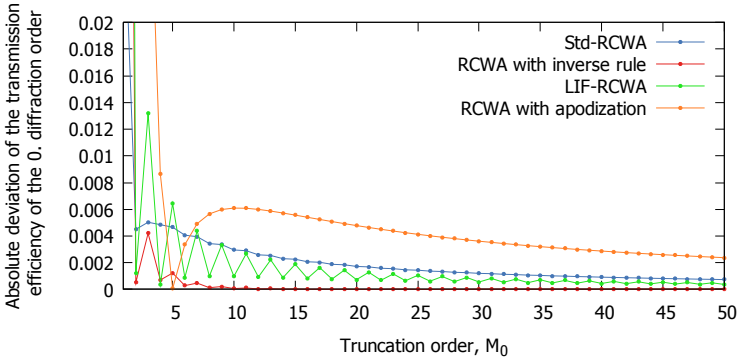
Another approach is to quantize the convergence of the fields themselves. This can be achieved by simply calculating the root mean square deviation (RMSD) between the fields of a low and a high mode calculation by

$$\mathrm{RMSD} = \sqrt{\frac{1}{N_x N_z} \sum_{k=0}^{N_z-1} \sum_{j=0}^{N_x-1} \left(w_{j,k} - \hat{w}_{j,k}\right)^2}, \tag{4.20}$$

where $N_x \times N_y$ is the number of sampling points of the field, $w_{j,k}$ is the calculated value at a position (x_j, z_k) in the field and $\hat{w}_{j,k}$ is the reference value of a fairly converged result (cf. Figs. 4.1c and 4.1d). In order to take all field components into account, the RMSD is calculated from intensity distributions $\left(\Rightarrow w_{j,k} = |\mathbf{E_1}\left(x_j, z_k\right)|^2\right)$.



(a) Convergence of the electric fields



(b) Convergence of the diffraction coefficients

Figure 4.4: Comparison between the convergence speed of (a) the electromagnetic near fields and (b) the diffraction coefficients. (a) shows the root mean square deviation between a high mode count ($M = 301$) field calculation with the standard RCWA and different field calculation methods with increasing mode count. (b) shows the corresponding convergence of the transmission efficiency of the 0th diffraction order. (The field calculation method of Lalanne and Jurek[67] is based on the RCWA with the inverse rule applied.)

Fig. 4.4a shows the RMSD between a high mode count calculation ($M = 301$) with the standard RCWA and different field calculation methods at lower but increasing mode counts. The fields were reconstructed in a region between $z = -0.45\,\mu m$ and $z = 0.75\,\mu m$ and $x = 0\,\mu m$ and $x = P_x = 0.6\,\mu m$ on a discrete grid with 1024×512 pixels.
The results are compared with the absolute deviation of the transmission efficiencies of the 0th diffraction order between an increasing mode count and the reference calculation with $M = 301$ modes in Fig. 4.4b.

4.2.3 Conclusion

Figs. 4.4a and 4.4b show that the convergence behavior of the fields is consistent with the convergence of the diffraction coefficients. This was expected, since the diffraction coefficients are calculated by using the continuous boundary conditions of the fields.
Apodization again shows the slowest convergence speed for the fields as well as for the diffraction coefficients. The convergence speed of the LIF-RCWA falls in between the standard RCWA and the RCWA with the inverse rule applied. The method does not rely on the inverse rule, but experiences oscillations in the convergence due to the lower sampling of the grating. The use of the inverse rule clearly achieves the fastest convergence rate for both the electromagnetic fields and the diffraction coefficients. Applying the method of Lalanne and Jurek[67] is even able to further improve the convergence speed of the fields. However, the method is only applicable to one-dimensional gratings and a study of the energy conservation in Sec. 5.3 will expose that the superior convergence of the inverse rule and the method by Lalanne and Jurek[67] does not come without a price.

Chapter 5

Absorption

When photodetectors, photovoltaic elements or photolithographic setups are designed, the main focus usually lies on the optimization of intensity distributions. They are simulated by scalar and rigorous diffraction theories in the time or modal domain, and although, these theories are sufficiently exact to predict intensities, they actually do not provide the relevant quantity that describes the performance of such devices. As a matter of fact, the true measurand is the amount of power that is absorbed in a given volume as it directly influences the electron generation in a detectors depletion region or the resist exposure of a lithographic setup. However, this requires a precise knowledge of the local absorption, which is not provided natively by most common diffraction theories.

This chapter will first discuss some simplified techniques to describe light absorption in classical optics and later proceed with a more general way to calculate the local absorption inside complex two- and three-dimensional structures using the standard RCWA according to Brenner[13] including an extension for the LIF-approach.

The ability to combine local absorption theory and the LIF-RCWA is indeed very powerful and expands the scope of rigorous design tools e.g. to optical storage technology, where focused beams in absorbing structures are used. Furthermore, at the end of this chapter, the theory of local absorption is used to test the different variants of the rigorous coupled-wave analysis with regard to their ability to conserve energy.

5.1 Lambert-Beer's law

A simple and often used method to calculate light absorption is described by the Lambert-Beer's law. It derives absorption in a homogeneous medium from a complex refractive index, which is irradiated by a monochromatic plane wave (cf. Figs. 5.1a and 5.1b).

According to Eq. 1.12, such plane wave, which propagates along z-axis ($\mathbf{s} = \mathbf{e_z}$) in an absorbing medium ($\hat{n} = n + i\kappa, \kappa \neq 0$)[1], can be described as

$$\begin{aligned}\mathbf{E}(\mathbf{r}) &= E_0\mathbf{E_1}\ \exp(i\mathbf{kr}) \\ &= E_0\mathbf{E_1}\ \exp(i(k_0\hat{n}\mathbf{sr})) \\ &= E_0\mathbf{E_1}\ \exp(i(k_0\hat{n}z)) \\ &= E_0\mathbf{E_1}\ \exp(i(k_0nz + ik_0\kappa z)) \\ &= E_0\mathbf{E_1}\ \exp(ik_0nz)\ \exp(-k_0\kappa z).\end{aligned} \tag{5.1}$$

Local intensity $I(z)$ is then determined by

$$\begin{aligned}I(z) &:= |\mathbf{E}|^2\ \exp(-2\kappa k_0 z) \\ &= I_0\, e^{-\alpha z},\end{aligned} \tag{5.2}$$

which shows an exponential decay along the penetration depth z. α is called the absorption coefficient and is wavelength dependent:

$$\alpha = 2\kappa k_0 = \frac{4\pi\kappa}{\lambda} \tag{5.3}$$

Comparing the intensity at a position z with the initial intensity at the position of the light source (at $z = 0$) results in the transmission magnitude at position z:

$$\mathcal{T}(z) = \frac{I(z)}{I(0)} \tag{5.4}$$

Hence, relative absorption up to z can be described by

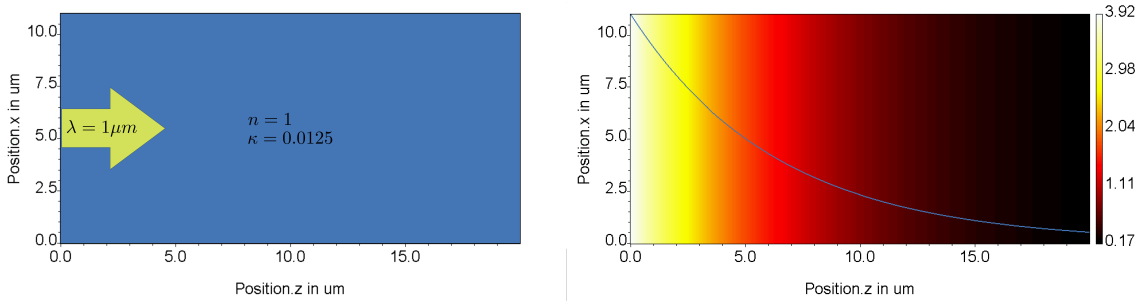
$$\mathcal{A}(z) = 1 - \mathcal{T}(z), \tag{5.5}$$

since the scope of this method is limited to media without any borders, where reflection must be accounted for. Finally, local absorption is simply the derivative along the penetration depth:

$$a(z) = \frac{\partial}{\partial z}\mathcal{A}(z) \tag{5.6}$$

[1]The relation between a complex refractive index $\hat{n}$ and $\hat{\epsilon}$ is $\hat{n} = \sqrt{\hat{\epsilon}\hat{\mu}}$. More details on this topic can be found in App. A.2.

Example and discussion



(a) Sketch of a homogeneous medium in which light is absorbed due to the imaginary part of the refractive index

(b) Local absorption (= rel. absorption/mm) in a homogeneous medium

Figure 5.1: Example of the use of Lambert-Beer's law for the calculation of local absorption in a homogeneous medium

Fig. 5.1b shows the local absorption of light at a wavelength of $\lambda = 1\,\mu m$ inside a homogeneous medium with $n = 1$ and $\kappa = 0.0125$ according to Lambert-Beer's law and Eq. 5.5. The results are in fact pretty accurate for this specific example. Nevertheless, the theory is very limited, because homogeneity must not only be ensured inside a considered object: Assuming this object is still homogeneous, but the real-part of its refractive index differs from the surrounding, then reflection would occur at all interfaces, which would not be considered by Lambert-Beer's law.

5.2 General theory of local absorption

In order to calculate local absorption for practical applications and scenarios, which are relevant for this work, like in the domain of diffraction gratings, a general absorption theory must also respect angular incidence and reflection and include multi layer systems and even structured layers.

5.2.1 Simple approach for layerwise absorption

In classical multi layer theory [21] as well as in modal methods (cf. Sec. 2.8), larger diffractive systems are often partitioned into layers with constant permittivity in the propagation direction of the light. Therefore, it is sometimes sufficient to know the amount of light, that is absorbed in a single layer l, like, for example, in the photoactive layer of a solar cell. A simple approach to obtain this information assumes, that the absorption in that specific layer l does not influence the absorbance of any other layer. In other words, changing only the absorption property in one layer, should not quantitatively change the absorbance of another layer.

$$\mathcal{A} := \sum_j A_j \tag{5.7a}$$

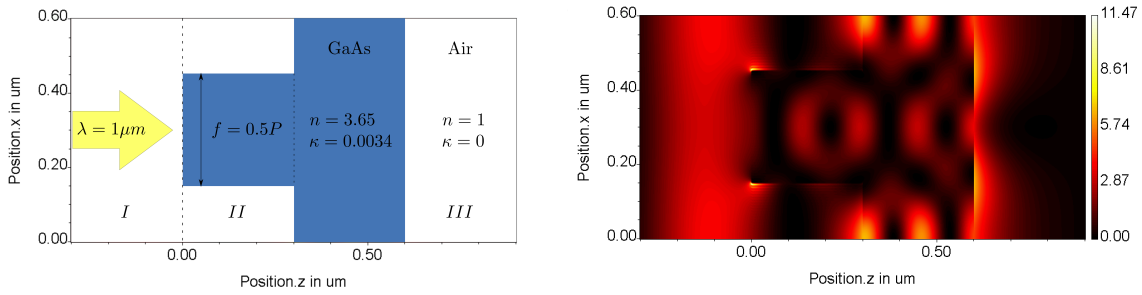
$$\mathcal{A}^{l;\bullet} := 1 - \mathcal{R}^{l;\bullet} - \mathcal{T}^{l;\bullet} \quad \text{with} \quad \kappa_l \neq 0 \Rightarrow A_l \neq 0 \tag{5.7b}$$

$$\mathcal{A}^{l;\circ} := 1 - \mathcal{R}^{l;\circ} - \mathcal{T}^{l;\circ} \quad \text{with} \quad \kappa_l = 0 \Rightarrow A_l = 0 \tag{5.7c}$$

$$A_l := \mathcal{A}^{l;\bullet} - \mathcal{A}^{l;\circ} \tag{5.7d}$$

Eq. 5.7a simply illustrates the fact, that the overall absorption ($\mathcal{A}$) of a grating is the sum of the absorption values A_j of all composing layers j. In order to obtain the absorption A_l of a single layer l, global transmission ($\mathcal{T}$) (cf. Eq. 2.71a) and global reflection ($\mathcal{R}$) (cf. Eq. 2.71b) of the entire structure is calculated twice, with $(l;\bullet)$ and without $(l;\circ)$ the absorption coefficients of the specific layer l. Each time a global absorption value is derived from the energy conservation law, similar to Eq. 5.5 in Lambert Beer's law, but this time taking reflection into account ($\mathcal{A} = 1 - \mathcal{R} - \mathcal{T}$, cf. Eqs. 5.7b and 5.7c). Finally, the difference between both global absorption values ($\mathcal{A}^{l;\bullet}$ and $\mathcal{A}^{l;\circ}$, cf. Eq. 5.7d) serves as an approximation for the absorption in layer l.

Example and discussion

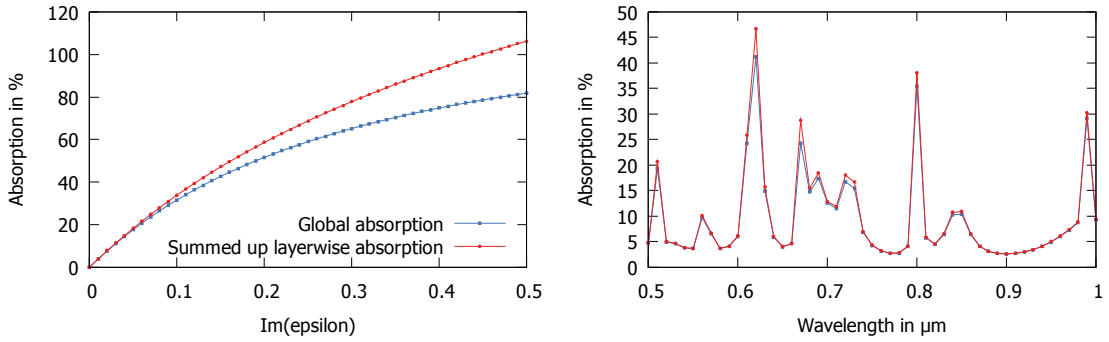


(a) Sketch

(b) Intensity distribution ($|E|^2$) with $M = 161$ Fourier modes

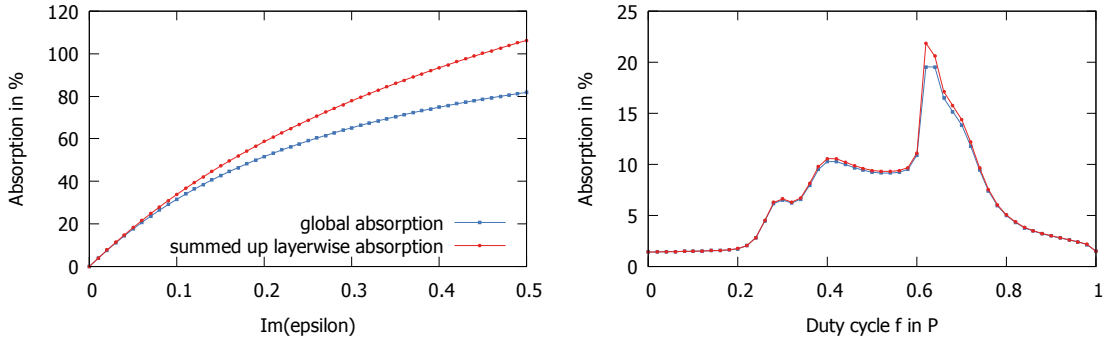
Figure 5.2: Example setup of a GaAs diffraction grating for testing the method of layerwise absorption

The diffraction grating depicted in Fig. 5.2a has a complex refractive index with $n = 3.65$ and $\kappa = 0.0034$ ($\Rightarrow$ GaAs) and consists of two layers with a thickness of $0.3\,\mu m$ each. The first layer has a duty cycle of $f = 0.5P$, whereas the second layer is homogeneous ($f = 1P$). The setup is illuminated with a TM polarized plane wave at a wavelength of $\lambda = 1\,\mu m$. Using the standard RCWA to obtain the diffraction efficiencies and the law of energy conservation to obtain a global absorption, one obtains a value of 9.20408 %. On the other hand, using the method of layerwise absorption one obtains the two separate absorption values 2.25681 % and 7.11633 % for the two individual layers, which sum up to a total of 9.37314 %. Thus, they do show a deviation of 1.83679 % from the expected global absorption.
Figs. 5.3a to 5.3d show, that the error primarily correlates with an increasing value of the imaginary part of the refractive index and the summed up layerwise absorption can even reach meaningless values of more the 100 %, whereas other parameters show only little and no systematic error.



(a) Total abs. for increasing values of the imaginary part of the refractive index

(b) Total absorption for increasing values of the wavelength



(c) Total absorption for increasing incident angles

(d) Total absorption for an increasing duty cycle of the first layer's grating

Figure 5.3: Influence of different parameters on the accuracy of the layerwise absorption method

The example shows that the method of layerwise disabling the absorption property can be used for a rough estimation of the absorption per layer. However, it's basic assumption, that the absorption property of the individual layers doesn't influence each other, is not correct. And besides the increasing inaccuracy for higher absorption coefficients, the method only provides absorption results for entire layers.

5.2.2 General local absorption in the standard RCWA

A more general approach was suggested by Brenner[13] and is derived here in detail as a supplement to the standard RCWA and the LIF-approach and extended to locally varying permeability. It is later used to demonstrate the potential of the LIF-RCWA and to test the different RCWA variants with regard on their ability to conserve energy.

For the exact calculation of the absorbed power in an arbitrary volume element, one has to go back to the fundamentals of electrodynamics. The conservation of energy states that the total energy of an isolated system is conserved over time. Thus all power that flows into the system is either reflected, transmitted or absorbed.

$$P_{\text{in}} = P_{\text{out}} + P_{\text{abs}} \tag{5.8}$$
$$\Rightarrow P_{\text{abs}} = P_{\text{in}} - P_{\text{out}} \tag{5.9}$$

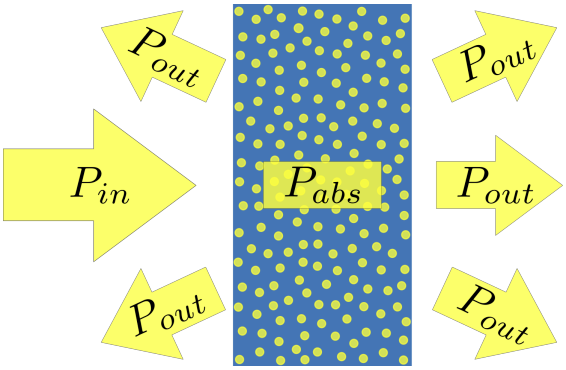


Figure 5.4: Illustration of the power flow through an absorbing structure

The Poynting vector $\mathbf{S} = \mathbf{E} \times \mathbf{H}$ describes the energy flux density of the steady state electromagnetic field measured in watt per square meter. Thus, the actual power can be derived by the divergence of the Poynting vector:

$$P_{\text{in}} := \int_V \nabla \mathbf{S_{in}}\, d^3V = -\oint_A \mathbf{S_{in}} d\mathbf{A} \tag{5.10a}$$
$$P_{\text{out}} := -\int_V \nabla \mathbf{S_{out}}\, d^3V = \oint_A \mathbf{S_{out}} d\mathbf{A} \tag{5.10b}$$
$$P_{\text{abs}} := -\int_V \nabla \mathbf{S}\, d^3V = -\oint_A \mathbf{S} d\mathbf{A} \tag{5.10c}$$

Since the origin of the incident light is a source of energy and of the electromagnetic field, the divergence of the Poynting vector of the incident is positive (cf. Eq. 5.10a). An absorbing

element or its infinite surrounding appears as an energy sink and thus the divergence of the corresponding Poynting vector is negative (cf. Eqs. 5.10b and 5.10c). According to Gauss's divergence theorem, the volume integral over the divergence of an energy flux can equally be described by a surface integral over the region's boundary as shown on the right in Eqs. 5.10a-5.10c. However, the theorem states, that the surface normals must always point outside the enclosed volume, which leads to a negative sign in Eq. 5.10a and 5.10c, where the flux is oriented towards the center of the enclosed volume. Thus positive energy is always ensured.

The consistency of the above definitions can be verified by substituting them back into Eq. 5.9:

$$P_{\text{abs}} = P_{\text{in}} - P_{\text{out}} \tag{5.11a}$$

$$\left(-\oint_A \mathbf{S} d\mathbf{A}_{\mathbf{near}}\right) = \left(-\oint_A \mathbf{S}_{\mathbf{in}} d\mathbf{A}_{\mathbf{far}}\right) - \oint_A \mathbf{S}_{\mathbf{out}} d\mathbf{A}_{\mathbf{far}} \tag{5.11b}$$

$$\oint_A \mathbf{S} d\mathbf{A}_{\mathbf{near}} = \oint_A \mathbf{S}_{\mathbf{in}} d\mathbf{A}_{\mathbf{far}} + \oint_A \mathbf{S}_{\mathbf{out}} d\mathbf{A}_{\mathbf{far}} \tag{5.11c}$$

$$\oint_A \mathbf{S} d\mathbf{A}_{\mathbf{near}} = \oint_A \mathbf{S} d\mathbf{A}_{\mathbf{far}}. \tag{5.11d}$$

The result shows that energy conservation holds regardless of whether the surface area is considered in the near or in far field. A similar consistency check can also be performed for the volume integrals.

In order to calculate the absorbed power, the divergence of the Poynting vector in Eq. 5.10c finally needs to be dissolved into electric and magnetic fields that are accessible by the RCWA:

$$\nabla \mathbf{S} = \nabla \left(\mathbf{E} \times \mathbf{H}\right) \tag{5.12}$$

Since only steady state systems are considered here, the time averaged Poynting vector is the essential quantity, and the physical monochromatic and harmonic field can be expressed by complex quantities (cf. A.3):

$$\nabla \langle \mathbf{S} \rangle = \frac{1}{2} \nabla \operatorname{Re}\left(\mathbf{E} \times \mathbf{H}^*\right) \tag{5.13a}$$

$$= \frac{1}{2} \operatorname{Re}\left(\left(\nabla \times \mathbf{E}\right) \mathbf{H}^*\right) - \frac{1}{2} \operatorname{Re}\left(\left(\nabla \times \mathbf{H}^*\right) \mathbf{E}\right) \tag{5.13b}$$

$$= -\frac{1}{2} \operatorname{Re}\left(\left(\mu\mu_0 \frac{\partial}{\partial t} \mathbf{H}\right) \mathbf{H}^* + \left(\mathbf{J}^* + \epsilon^* \epsilon_0 \frac{\partial}{\partial t} \mathbf{E}^*\right) \mathbf{E}\right) \tag{5.13c}$$

$$= -\frac{1}{2} \operatorname{Re}\left(\mu\mu_0 \frac{\partial \mathbf{H}}{\partial t} \mathbf{H}^* + \mathbf{J}^* \mathbf{E} + \epsilon^* \epsilon_0 \frac{\partial \mathbf{E}^*}{\partial t} \mathbf{E}\right) \tag{5.13d}$$

Here, $\mathbf{J}$ denotes the electric current density, which is neglectable for optical frequencies (cf. Sec. 1.1.4).

With the Ampère's circuital law (cf. Eq. 1.1a) and Maxwell–Faraday's equation (cf. Eq. 1.1b) applied one obtains

$$\nabla\langle\mathbf{S}\rangle = -\frac{1}{2}\,\mathrm{Re}\left(\epsilon^*\epsilon_0\frac{\partial\mathbf{E}^*}{\partial t}\mathbf{E} + \mu\mu_0\frac{\partial\mathbf{H}}{\partial t}\mathbf{H}^*\right). \tag{5.14}$$

Treating the permittivity as a complex quantity ($\epsilon = \epsilon_r + i\epsilon_i$) and the relative permeability as neutral ($\mu = 1$) and also resolving the real part operator by $\mathrm{Re}\,(z) = 1/2\,(z + z^*)$, the equation transforms to

$$\nabla\langle\mathbf{S}\rangle = -\frac{1}{4}\left(\epsilon_r\epsilon_0\,\left(\frac{\partial\mathbf{E}^*}{\partial t}\mathbf{E} + \frac{\partial\mathbf{E}}{\partial t}\mathbf{E}^*\right) + \mu_0\,\left(\frac{\partial\mathbf{H}}{\partial t}\mathbf{H}^* + \frac{\partial\mathbf{H}^*}{\partial t}\mathbf{H}\right) + i\epsilon_i\epsilon_0\,\left(\frac{\partial\mathbf{E}}{\partial t}\mathbf{E}^* - \frac{\partial\mathbf{E}^*}{\partial t}\mathbf{E}\right)\right), \tag{5.15}$$

in which the first two terms cancel out after the time derivation of the electric and magnetic fields as they are assumed to be monochromatic and harmonic
($\sim exp(-i\omega t) \Rightarrow \mathbf{F}\,(t) = \mathbf{F}e^{-i\omega t}$). This leads to

$$\nabla\langle\mathbf{S}\rangle = -\frac{\epsilon_i\epsilon_0}{2}\omega|\mathbf{E}(\mathbf{r})|^2. \tag{5.16}$$

Substituting Eq. 5.16 back into Eq. 5.10c and replacing $\epsilon_0\omega$ by k_0/Z_0 provides a first result for the absorbed power in a volume V:

$$P_{\mathrm{abs}} = -\int_V \nabla\langle\mathbf{S}\rangle\,d^3V = \frac{k_0}{2Z_0}\int_V \mathrm{Im}\,(\epsilon\,(\mathbf{r}))\,|\mathbf{E}(\mathbf{r})|^2\,d^3V \tag{5.17}$$

The electric field $\mathbf{E}(\mathbf{r})$ can now be separated into a product of constant amplitude E_0 and the system's response $\mathbf{E_1}(\mathbf{r})$ to an incident field with amplitude $E = 1$ ($\Rightarrow \mathbf{E}(\mathbf{r}) = E_0\mathbf{E_1}(\mathbf{r})$), leading to

$$P_{\mathrm{abs}} = \frac{k_0}{2Z_0}|E_0|^2\int_V \mathrm{Im}\,(\epsilon\,(\mathbf{r}))\,|\mathbf{E_1}(\mathbf{r})|^2\,d^3V. \tag{5.18}$$

The incident power P_{in} on the other hand can be derived from the surface integral in Eq. 5.10a:

$$P_{in} = -\oint_A \langle \mathbf{S_{in}} \rangle d\mathbf{A} \tag{5.19a}$$

$$= -\iint_A \frac{1}{2} \operatorname{Re}\left(\mathbf{E} \times \mathbf{H}^*\right) d\mathbf{A} \tag{5.19b}$$

$$= -\iint_A \frac{1}{2Z_0 k_0} |\mathbf{E}|^2 \operatorname{Re}\left(\mathbf{k_i}\right) d\mathbf{A} \tag{5.19c}$$

$$= -\iint_A \frac{1}{2Z_0 k_0} |E_0|^2 \operatorname{Re}\left(\mathbf{k_i}\, \mathbf{n_A}\right) dA \tag{5.19d}$$

$$= \frac{1}{2Z_0 k_0} \operatorname{Re}\left(k_{i;z}\right) |E_0|^2 \iint_A dA \tag{5.19e}$$

$$= \frac{A}{2Z_0 k_0} \operatorname{Re}\left(k_{i;z}\right) |E_0|^2 \tag{5.19f}$$

Because in the present setup (cf. Fig. 5.4) the incident light is pointing from negative z on to the grating and the grating itself is periodically infinite in x- and y-direction, the closed integral can be replaced by a surface integral over the area $A = P_x P_y$, which lays in the x/y-plane somewhere in region I at $z \leq 0$. $\mathbf{n_A} = -\mathbf{e_z}$ is its normal unit vector pointing back in region I, whereas $\mathbf{k_i}$ is the wave vector in the direction of the field propagation and points in positive z-direction. Thus, its product is $-k_{i;z}$. The real part operator $\operatorname{Re}(\cdot)$ indicates that only a propagative incident wave carries energy.

Integrated absorption

The relation between the absorbed power and the incident power finally leads to an equation for the relative absorbed power in a volume V, independent of the unknown constant E_0:

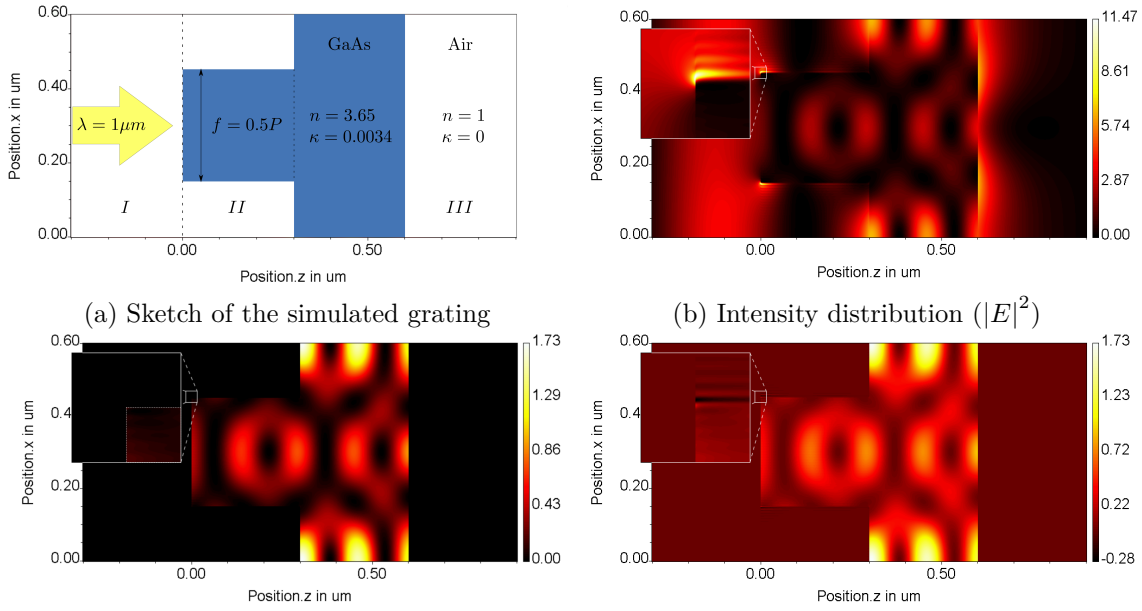
$$\text{Abs} = \frac{P_{\text{abs}}}{P_{\text{in}}} = \frac{1}{A} \frac{k_0^2}{k_{i;z}} \int_V \operatorname{Im}\left(\epsilon\left(\mathbf{r}\right)\right) |\mathbf{E_1}(\mathbf{r})|^2 \, d^3V \tag{5.20}$$

This shall be referred to as the integrated absorption.

With $d^3V = dx\, dy\, dz$, $A = P_x P_y$ and $dq = P_q / N_q$ for $q = x, y$, Eq. 5.20 can be expressed in a discretized form as

$$\text{Abs} = \frac{dz}{N_x \cdot N_y} \frac{k_0^2}{k_{i;z}} \sum_j^{N_x - 1} \sum_k^{N_y - 1} \sum_l^{N_z - 1} \operatorname{Im}\left(\epsilon(r_{j,k,l})\right) |\mathbf{E_1}(r_{j,k,l})|^2 \,. \tag{5.21}$$

Example and discussion



(a) Sketch of the simulated grating

(b) Intensity distribution ($|E|^2$)

(c) Local absorption from *design epsilon*

(d) Local absorption from *reconstructed epsilon*

Figure 5.5: Calculation of local absorption in a GaAs-grating on a 1024×512-sized grid with $M = 161$ Fourier modes.

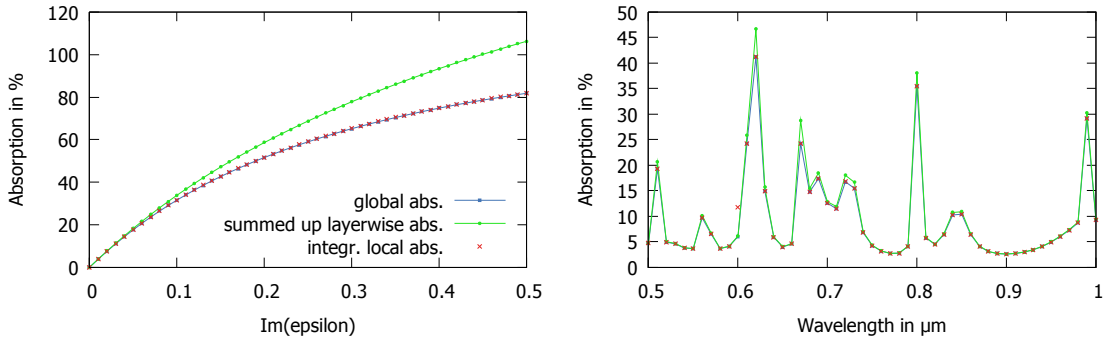
The method of layerwise absorption in Sec. 5.2.1 only allows a rough estimation of the total absorption per layer. By contrast, the theory of local absorption is able to provide local absorption with arbitrary spatial resolution[2]. Fig. 5.5a shows the same setup as in the last example. Local absorption is now calculated on a grid with 1024×512 discrete spatial points and $M = 161$ Fourier modes (Std-RCWA) in two different ways: In Fig. 5.5c the imaginary part of the permittivity is taken straight from $\epsilon(x)$, exactly as it was designed and defined in Fig. 5.5a. In this case the permittivity is not mode-limited and as such further referred to as *design epsilon*. In Fig. 5.5d the mode-limited permittivity $\epsilon^{(M)}(x)$ is used, further referred to as *reconstructed epsilon* (cf. Eq. 2.37). Since the RCWA is a modal method and already processes the permittivity as a mode-limited quantity, the latter approach is the more consistent one and used as such in all subsequent local absorption calculations. Nevertheless, the integrated absorption of both method provides very similar results and compared to the layerwise absorption approach they both show superior consistency with the global absorption reference that is derived from the diffraction efficiencies (cf. Eq. 5.7b). This is true even though the reconstructed epsilon absorption shows smoother edges in lateral direction and additional over- and undershoots due to the mode limitation (cf. Fig. 5.5d). Tab. 5.1 shows the comparison between the global absorption and the absorption results of the different methods for the example in Fig. 5.5a.

[2]Of course, if the chosen spatial resolution lies above the modal resolution used in the calculation, the electric field is only Fourier interpolated and so the calculated local absorption is only an approximation.

Figs. 5.6a to 5.6d also illustrates a high agreement between the integrated local absorption and the global absorption for a diversity of parameter variations including an increasing absorption property. In fact the result match to such degree, that it appears safe to assume, that the remaining errors solely originate from the sampling of the electric field. Unfortunately, this intuition is misguided, even though the problem doesn't lie in the theory of local absorption. The topic is further discussed in Sec. 5.3.

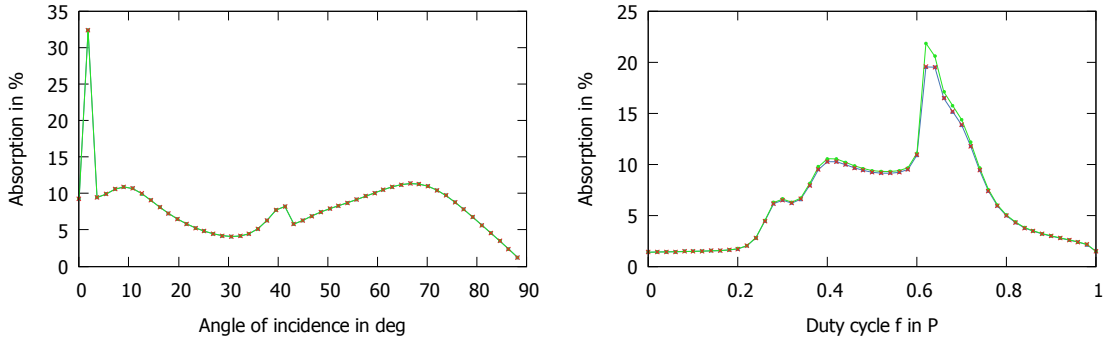
method	abs. in layer 1	abs. layer 2	$\sum$	deviation
layerwise absorption	2.25681%	7.11633%	9.37314%	1.83679%
local abs. (dsgn. eps)	2.18357%	7.03371%	9.21728%	0.14341%
local abs. (rec. eps)	2.18215%	7.03371%	9.21586%	0.12799%
global absorption			**9.20408%**	

Table 5.1: Comparison between different methods to calculate absorption for the example grating in Fig. 5.5a and their deviation from the global absorption.



(a) Total absorption for increasing values of the imaginary part of the refractive index

(b) Total absorption for increasing values of the wavelength



(c) Total absorption for increasing incident angles

(d) Total absorption for an increasing duty cycle of the first layer's grating

Figure 5.6: Influence of different parameters on the accuracy of the local absorption theory. Calculations were performed on a 1024×512-sized grid with $M = 81$ Fourier modes.

5.2.3 Complex local permeability

If also the permeability μ is considered to be a complex and locally varying material property of the grating, the same approach as described in the previous section lead to an even more general expression for absolute local absorption that is

$$P_{\text{abs}} = \frac{k_0}{2} \int_V \left(\frac{1}{Z_0} \operatorname{Im}\left(\epsilon\left(\mathbf{r}\right)\right) |\mathbf{E}(\mathbf{r})|^2 + Z_0 \operatorname{Im}\left(\mu\left(\mathbf{r}\right)\right) |\mathbf{H}(\mathbf{r})|^2 \right) d^3V. \tag{5.22}$$

However, complex permeability at optical frequencies are usually only expected for photonic metamaterial. In the surrounding medium, the permeability is, therefore, assumed to be a constant, real number μ_I and thus the irradiated power is now defined as

$$P_{in} = \frac{A}{2Z_0 k_0 \mu_I} \operatorname{Re}\left(k_{i;z}\right) |E_0|^2. \tag{5.23}$$

5.2.4 General local absorption in the LIF-RCWA

For the LIF-RCWA, local absorption is calculated very similar. Only the incoming power

$$P_{\text{in}}^{\text{LIF}} = -\oint_A \frac{1}{2} \operatorname{Re}\left(\mathbf{E} \times \mathbf{H}^*\right) d\mathbf{A}, \tag{5.24}$$

is now composed of a superposition of multiple electric and magnetic plane waves $\mathbf{L}_{m,n}^E$ and $\mathbf{L}_{m,n}^H$. Again, with A as the rectangular incidence area of size $P_x P_y$ and the z-axis as it's surface normal $\mathbf{n_A} = -\mathbf{e}_z$ the closed integral can be replaced by a double integral over the two grating periods and the vector product of electric and the magnetic field only needs to be evaluated for the z-component:

$$P_{\text{in}}^{\text{LIF}} = \oint_A \frac{1}{2} \operatorname{Re}\left(E_x H_y^* - E_y H_x^*\right) dA \tag{5.25a}$$

$$= \frac{1}{2} \int_0^{P_x} \int_0^{P_y} \operatorname{Re}\left(F\right) dxdy \tag{5.25b}$$

with

$$F = \sum_{m,n} \sum_{m',n'} \left(L_{x;m,n}^E L_{y;m',n'}^{*;H} - L_{y;m,n}^E L_{x;m',n'}^{*;H} \right) e^{2\pi i \left(\frac{m-m'}{P_x} x + \frac{n-n'}{P_y} y \right)} e^{i\gamma_{I;m,n} z} \tag{5.26}$$

The z coordinate can be chosen arbitrarily in region I. Hence, choosing $z = 0$ will let the last term in Eq. 5.25b disappear. The integrals can now be solved as

$$P_{\text{in}}^{\text{LIF}} = \frac{1}{2}\,\text{Re}\left(\sum_{m,n}\sum_{m',n'}\left(L^{E}_{x;m,n}L^{*;H}_{y;m',n'} - L^{E}_{y;m,n}L^{*;H}_{x;m',n'}\right)\right) P_x\delta_{m',m}P_y\delta_{n',n} \tag{5.27a}$$

$$= \frac{1}{2}P_xP_y\sum_{m,n}\text{Re}\left(L^{E}_{x;m,n}L^{*;H}_{y;m,n} - L^{E}_{y;m,n}L^{*;H}_{x;m,n}\right). \tag{5.27b}$$

Finally the magnetic components can be expressed by electric components through the conversion matrix $\mathbf{C}$ (cf. Eq. 2.28) with

$$\mathbf{L}^{H}_{\perp} = \mathbf{C}\,\mathbf{L}^{E}_{\perp} = \begin{pmatrix}\mathbf{C}_{xx} & \mathbf{C}_{xy}\\ \mathbf{C}_{yx} & \mathbf{C}_{yy}\end{pmatrix}\begin{pmatrix}\mathbf{L}^{E}_{x}\\ \mathbf{L}^{E}_{y}\end{pmatrix} \tag{5.28}$$

leading to

$$P_{\text{in}}^{\text{LIF}} = \frac{A}{2}\sum_{m,n}\text{Re}\left(L^{E}_{x;m,n}\left(C_{yx;m,n}L^{E}_{x;m,n} + C_{yy;m,n}L^{E}_{y;m,n}\right)^{*}\right. \tag{5.29a}$$

$$\left. - L^{E}_{y;m,n}\left(C_{xx;m,n}L^{E}_{x;m,n} + C_{xy;m,n}L^{E}_{y;m,n}\right)^{*}\right) \tag{5.29b}$$

$$= \frac{A}{2}\,\text{Re}\left(\left[\mathbf{C}\mathbf{L}^{\mathbf{E}}_{\perp}\right]^{*}\left[\mathbf{J}\mathbf{L}^{\mathbf{E}}_{\perp}\right]\right) \quad \text{with} \quad \mathbf{J} = \begin{pmatrix}0 & -\mathbf{I}\\ \mathbf{I} & 0\end{pmatrix} \tag{5.29c}$$

2D-Case simplification

In case of only one-dimensional gratings in classical mounting (cf. Fig. 1.3a), $\mathbf{k}_y$ disappears and the diffraction problem can be separated into two cases of polarization, namely TE and TM, which allows the following simplifications of Eq. 5.29b:

TE polarization:
$E_x = E_z = H_y = 0 \Rightarrow L^{E}_{x;m,n} = L^{E}_{z;m,n} = L^{H}_{y;m,n} = 0$

$$\Rightarrow P_{\text{in}}^{\text{LIF}} = \frac{A}{2Z_0k_0\mu_I}\sum_{m,n}\text{Re}\left(k_{z;m,n}\right)\left|L^{E}_{y;m,n}\right|^{2}. \tag{5.30}$$

TM polarization:
$E_y = H_x = H_z = 0 \Rightarrow L^E_{y;m,n} = L^H_{x;m,n} = L^H_{z;m,n} = 0$

$$\Rightarrow P_{\text{in}}^{\text{LIF}} = \frac{A}{2Z_0 k_0 \mu_I} \sum_{m,n} \text{Re}\left(k_{z;m,n}\right) \left|L^E_{x;m,n}\right|^2. \tag{5.31}$$

Both cases show a similar form as the derived incident power in the standard case (cf. Eq. 5.19f).

The normalized integrated absorption is again given by

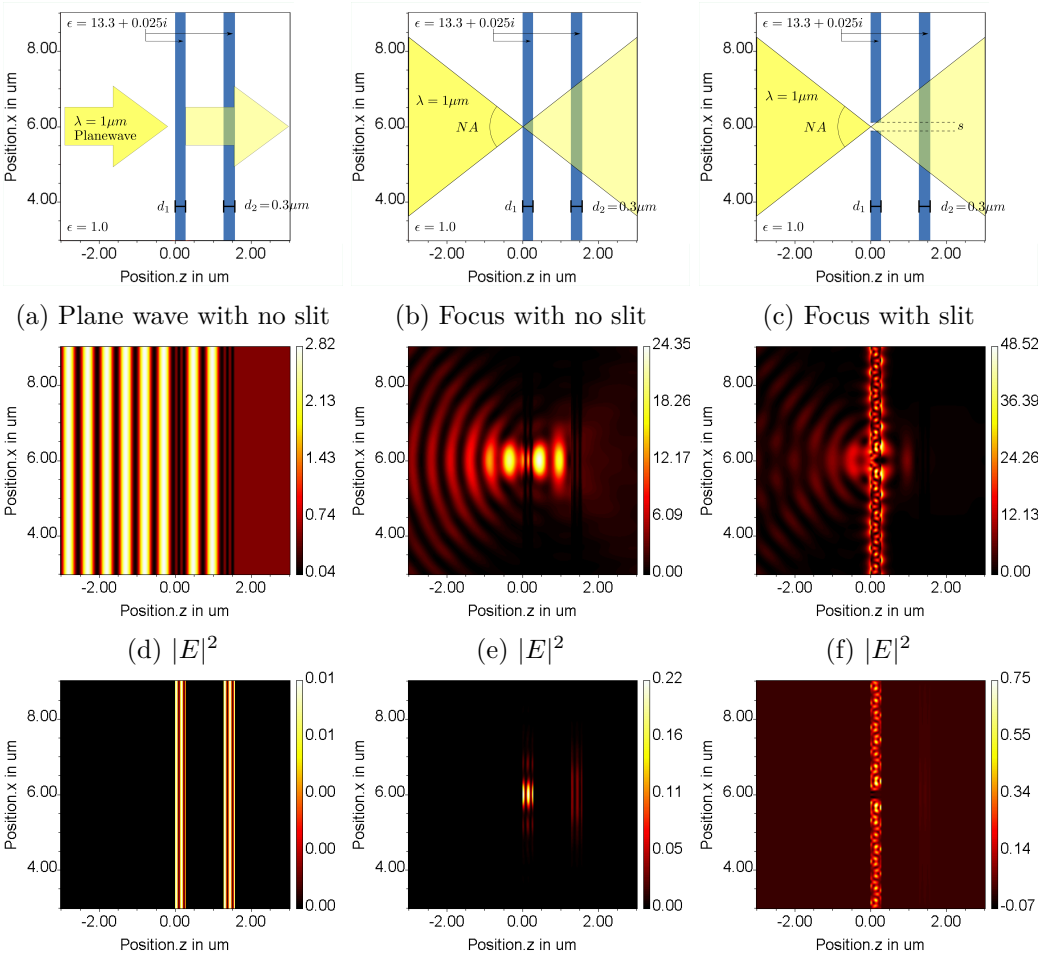
$$\text{Abs}^{\text{LIF}}\left(\mathbf{r}\right) = \frac{P_{\text{abs}}}{P_{\text{in}}^{\text{LIF}}} \quad \text{with} \tag{5.32a}$$

$$P_{\text{abs}} = \frac{k_0}{2Z_0} \int_V \text{Im}\left(\epsilon\left(\mathbf{r}\right)\right) \left|\mathbf{E}(\mathbf{r})\right|^2 d^3V. \tag{5.32b}$$

Example and discussion

The following example consists of a simple setup with two absorbing layers of GaAs, which are separated by a gap of air (distance $= 1\,\mu m$). The goal was to optimize the setup parameters in order maximize the absorbed power in layer 1. The first column of Fig. 5.7 shows the setup illuminated by a TM polarized plane wave with wavelength $\lambda = 1\,\mu m$. With a thickness of $d_1 = \lambda/n = 0.2742\,\mu m$ and $d_2 = 0.3\,\mu m$ the two layers only absorb 1.18741 % and 1.35813 % of the light respectively. In setup 2 (cf. second column of Fig. 5.7) the plane wave is replaced by a focused beam (as described in Sec. 3.10.3) with a numerical aperture of $NA = 1$. This causes layer 1 to absorb twice the power, while layer 2 remains unchanged. Finally, in setup 3 (cf. third column of Fig. 5.7), the first layer is interrupted by a small slit of $s = 0.23881\,\mu m$ and the absorption in this layer suddenly reaches 53.80354 %, while at the same time the absorption in layer 2 drops down to 0.60943 %. This indicates that, although the slit is not even a third of the wavelength in size it now collects most of the light and channels it into layer 1, where it remains contained due to total reflection at both of its interfaces. Just by moving both, slit and focal spot, to a different position, it is also easily possible to optimize for maximum absorption in a hidden layer, like layer 2 in this case.

The example shows, that in combination the LIF-RCWA and the theory of local absorption can be used to improve the performance of devices, where high absorptions in confined locations are the critical parameter, for example in the active regions of a solar cell or a photodetector. The theory can in fact also be used to optimize for absorption in even smaller regions inside a layer.



(a) Plane wave with no slit (b) Focus with no slit (c) Focus with slit

(d) $|E|^2$ (e) $|E|^2$ (f) $|E|^2$

(g) Local absorption (h) Local absorption (i) Local absorption

Figure 5.7: Comparison of three configurations with regard to their absorption performance in layer 1. All calculations were performed with $M = 201$ Fourier modes.

	gl. trans.	gl. refl.	abs. in layer 1	abs. in layer 2	gl. abs.
setup 1	50.89931%	46.56045%	1.18741%	1.35813%	2.54024%
setup 2	62.96804%	33.17753%	2.39485%	1.43589%	3.85443%
setup 3	25.67581%	20.24959%	**53.45612**%	0.60617%	54.07461%

Table 5.2: Transmission, reflection and absorption value for the three example configurations in Fig. 5.7

5.3 Energy conservation in the RCWA

5.3.1 Energy conservation in absorbing gratings

The theory of local absorption can not only be used for the optimization of optical gratings, but also to verify the RCWA itself.

Moharam and Gaylord claim that *conservation of energy for a lossless grating [...] should be achieved to an accuracy of at least 1 part in 1^{10}, regardless of the number of terms in the field-expansion series that are retained in the formulation*[91]. And in fact, this statement is widely accepted. So again, for a lossless grating the power deficit is defined as the difference between the irradiated power and the sum of reflected and transmitted power (cf. Eq. 2.71)

$$\text{power deficit} := 1 - \mathcal{R} - \mathcal{T} \tag{5.33}$$

should effectively be zero. For an absorbing grating on the other hand, it seems natural to simply address the power deficit to the absorption in the grating as done in Sec. 5.2.2 (cf. Example and Discussion). However, the situation is not so clear. In fact, Russel[109] already criticized the RCWA's truncation scheme to be the cause of a violation of the law of energy conservation even for lossless gratings. So, for an absorbing grating the question of energy conservation becomes even more significant, since there is no way to distinguish between a numerically induced power deficit (cf. Eq. 5.33) and the actually absorbed power in the grating. At this point the concept of local absorption offers a way for clarification, as it provides a different way of calculating the absorbed power (cf. Eq. 5.21):

$$\text{integrated absorption} := \frac{dz}{N_x\, N_y} \frac{k_0^2}{k_{i;z}} \sum_{j}^{N_x-1} \sum_{k}^{N_y-1} \sum_{l}^{N_z-1} \operatorname{Im}\left(\epsilon(r_{j,k,l})\right) \left|\mathbf{E_1}(r_{j,k,l})\right|^2 \tag{5.34}$$

Since the two definitions for the power deficit (cf. Eq. 5.33) and the integrated absorption (cf. Eq. 5.34) are normalized in terms of the irradiated power, they can be directly compared and any deviation will indicate a violation of energy conservation either in the RCWA algorithm itself or in the calculation of the electromagnetic fields. In other words, if energy is fully conserved, the following equation should always hold:

$$\text{integrated absorption} \stackrel{!}{=} \text{power deficit} \tag{5.35}$$

So the following goal is primarily to look for consistency or discrepancies of Eq. 5.35 for the different RCWA variants and field calculation methods, which have been discussed so far.

For this purpose the same two-layered GaAs grating is analyzed as in the previous examples (cf. Fig. 5.5a). In order to minimize the influence of sampling errors, the electric fields, which are needed for the calculation of the local absorption (cf. Eq. 5.34), are reconstructed on N_x grid points in lateral direction from $x = 0\,\mu m$ to $x = P_x = 0.6\,\mu m$. N_x is the number of sampling points, which was already used for the discrete Fourier expansion of the permittivity. In case of the LIF-RCWA N_x is predetermined by the fixed relation between the sampling points in real space and in the Fourier domain: $N_x = 2M_0 + 1 = M$. For every other RCWA-variant there is no single number of Fourier coefficients that can be associated with the expansion of the permittivity (cf. Sec. 3.5), and the common understanding is that a larger number of sampling points leads to less reconstruction artifacts. So, for all methods except the LIF-RCWA, N_x was chosen to be 512. Since the RCWA does not use a Fourier expansion in the direction of z, there is also no ideal number of reconstruction points N_z in any of the methods and again a larger number is assumed to lead to a higher accuracy of the fields. The absorption of the grating is determined by the discretized integral in Eq. 5.34, where every point $r_{j,k,l}$ marks the center of a $dx \times dy \times dz$ cubic volume element. While the lateral displacement of the cubic elements is arbitrary due to the periodicity of the grating, one has to ensure a perfect alignment of the cubic integration elements in the grating region in the direction of z (cf. Fig. 5.8) in order to minimize sampling errors. With a sampling of $N_z = 1024$ over the distance $z = -0.01\,\mu m$ to $z = 0.61\,\mu m$, in which the grating reaches from $z = 0\,\mu m$ to $z = d = 0.6\,\mu m$ the correct alignment is ensured.

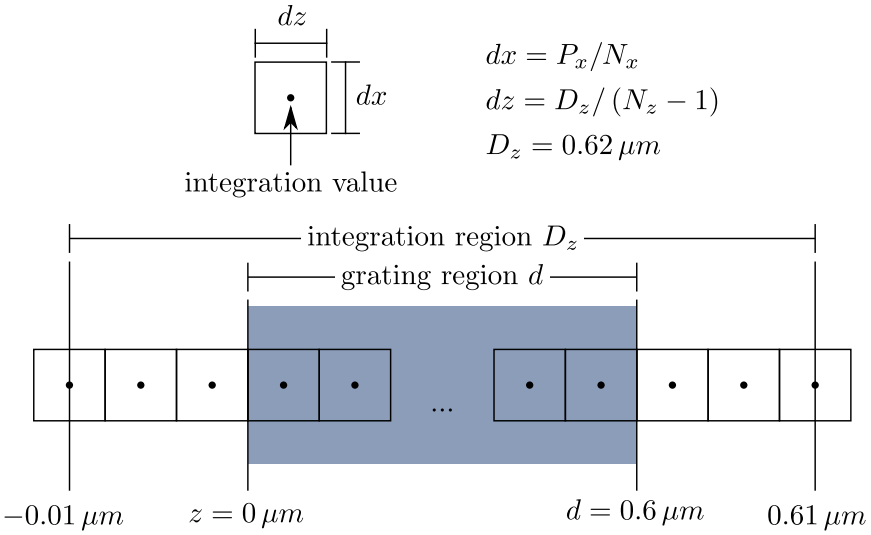
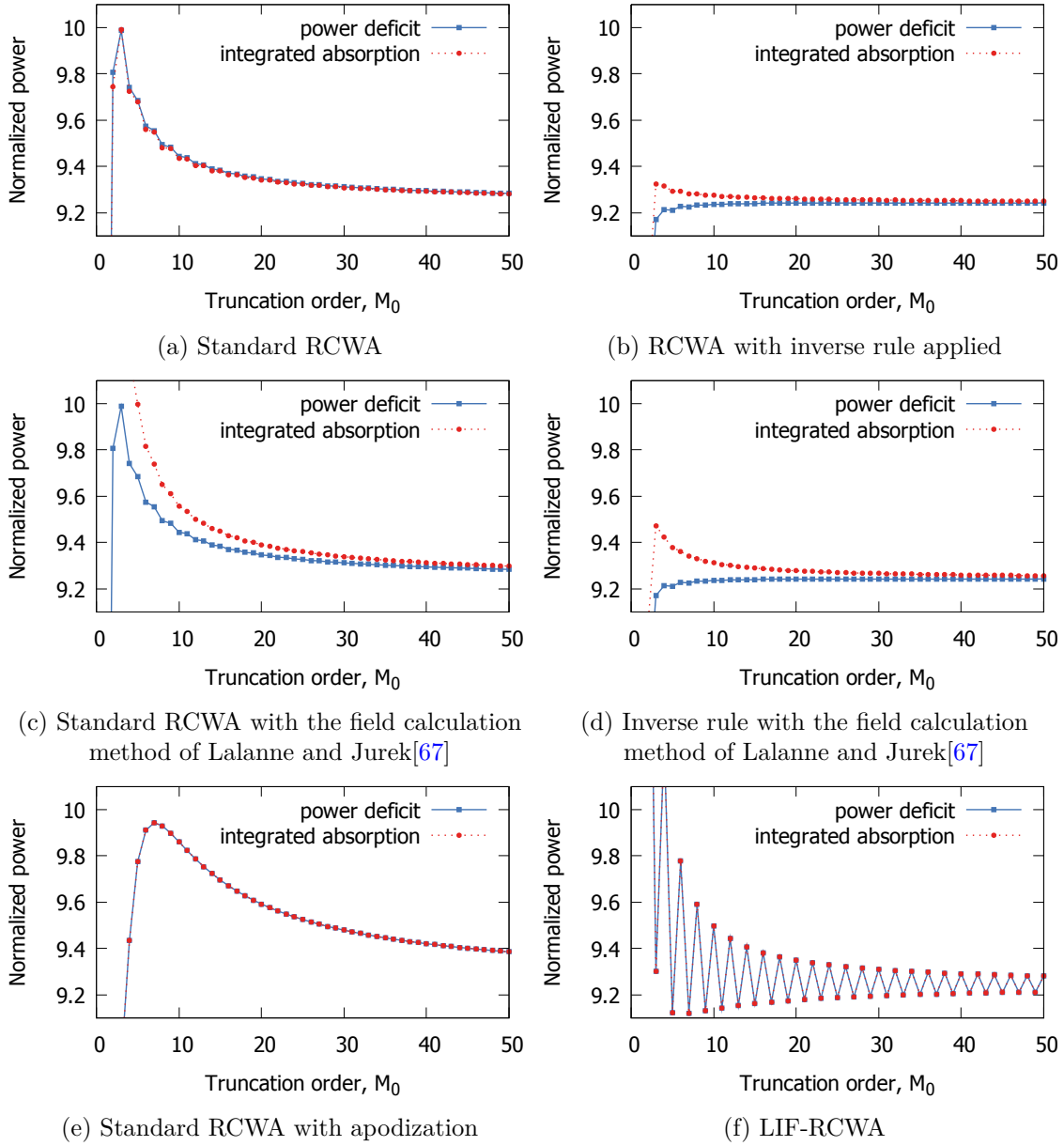


Figure 5.8: Schematic of a simple integration scheme

Figs. 5.9a to 5.9f show the converging values for the power deficit and the integrated absorption of the different RCWA variants. And in fact, deviations, which indicate a violation of the energy conservation, can be found in almost every method, especially at a low order of truncation.

Interestingly, this phenomenon is already noticeable for the standard RCWA (cf. Fig. 5.9a). It becomes even more apparent, if the inverse rule (cf. Fig. 5.9b) is applied and further amplifies, if the field calculation method of Lalanne and Jurek[67] (cf. Figs. 5.9c and 5.9d) is used. By contrast, applying apodization on the permittivity function seems to eliminate almost any difference (cf. Fig. 5.9e). Finally, the LIF-RCWA shows no deviation between power deficit and integrated absorption up to five decimal places (cf. Fig. 5.9f). The raw deviations are again depicted in Fig. 5.10 and can be reviewed individually for the first five orders of truncation in Tab. 5.3.



(a) Standard RCWA

(b) RCWA with inverse rule applied

(c) Standard RCWA with the field calculation method of Lalanne and Jurek[67]

(d) Inverse rule with the field calculation method of Lalanne and Jurek[67]

(e) Standard RCWA with apodization

(f) LIF-RCWA

Figure 5.9: Comparison between the power deficit and the integrated absorption for different RCWA variants.

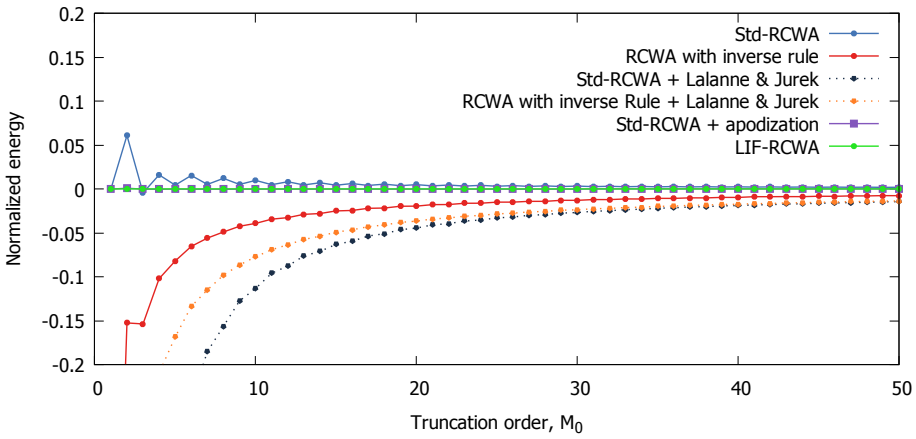


Figure 5.10: Absolute deviation between the power deficit and the integrated absorption for all presented methods

Method / M_0	1	2	3	4	5	6
(a)	0.00000	0.06103	−0.004214	0.01617	0.00450	0.01552
(b)	−0.69631	−0.15221	−0.15335	−0.10187	−0.08173	−0.06511
(c)	−2.06141	−1.41332	−0.74358	−0.46371	−0.31362	−0.24104
(d)	−0.95983	−0.36502	−0.30170	−0.20918	−0.16815	−0.13335
(e)	−0.00000	0.00100	−0.00001	−0.00016	−0.00002	−0.00013
(f)	−0.00000	0.00001	0.00000	0.00000	0.00000	0.00000

Table 5.3: Absolute deviation between the power deficit and the integrated absorption for different RCWA variants at $M_0 = 0$ to 5 Fourier modes: (a) Std-RCWA, (b) RCWA with inverse rule, (c) Std-RCWA + Lalanne and Jurek[67], (d) RCWA with inverse rule + Lalanne and Jurek[67], (e) Std-RCWA with apodization and (f) LIF-RCWA

Interpretation of the results

There are several factors that may influence either the power deficit, the integrated absorption or both:

Incorrect sampling

The correct sampling of the absorption integral in Eq. 5.34 is in fact crucial and an obvious reason for a discrepancy in Eq. 5.35. Fig. 5.11 illustrates how the sampling influences the power deviation. A finer sampling in the direction of z reduces the deviation for both the standard RCWA and the LIF-RCWA (cf. Fig. 5.11b and 5.11d). But in case of the standard RCWA, the deviation does not seem to converge towards zero, but maintains a residual value. In lateral direction the LIF-RCWA reveals an optimum sampling as predicted at $N_x = 9$ (sampling of the permittivity: $N_x = 2M_0 + 1 = M$, here $M_0 = 4 \Rightarrow N_x = 9$), where the deviation effectively disappears (cf. Fig. 5.11c). In the standard RCWA, however, the deviation remains for any sampling (cf. Fig. 5.11a).

Non-cyclic Toeplitz matrix

In Sec. 3.5 it was already pointed out, that the standard RCWA uses an inconsistent truncation scheme in the bandwidth limiting process of the grating permittivity. This is substantiated by the fact, that the LIF-RCWA has an optimum lateral sampling at $N_x = 2M_0 + 1 = M$, whereas no single number of Fourier modes can be associated with the permittivity matrix $[\![\epsilon]\!]$. Since the truncation inconsistency especially concerns the higher Fourier modes in the outer diagonals of the Toeplitz matrix $[\![\epsilon]\!]$, it appears reasonable that a suppression of such higher modes by a suitable apodization function will reduce its symptoms. And in fact this exact phenomenon can be observed in Fig. 5.9e and Tab. 5.3, where the deviation almost disappears, when using apodization. Moreover, it was demonstrated in Sec. 2.6 that apodization does not fully cure the Gibbs phenomenon as the irregularity of the displacement field at the concurrent jump discontinuity remains. This in fact indicates that the violation of energy conservation is not related to this phenomenon, but can be attributed to the erroneous truncation scheme.

Inverse rule

It is known that the inverse rule is only an approximation when finite, non-cyclic Toeplitz matrices are used and that again the approximation becomes better with an increasing order of truncation. Since the inverse matrix $[\![1/\epsilon]\!]^{-1}$ uses the same inconsistent truncation scheme as the standard RCWA, there is a strong indication, that energy is also not conserved as the deviation increases significantly - despite the fact that the rule improves the convergence of the RCWA algorithms. The distinction between the algorithms of the LIF-RCWA and the standard RCWA, when applied on a single plane incident wave is exactly the use of a consistent truncation scheme in the LIF-RCWA, which leads to a disappearance of the deviation.

Mixing band-limited and non-band-limited quantities

The usage of a non-band-limited permittivity function in an otherwise bandwidth-limited calculation, like it is proposed in the field calculation method of Lalanne and Jurek[67], obviously affects the energy balance. The reason is that the non-bandwidth-limited permittivity redistributes the power of the incident wave over infinite frequencies. But since only a finite subset of the frequencies is considered in the further analysis, energy is lost.

By avoiding all the sources of errors of the above and especially by reducing the number of lateral reconstruction points N_x to the number of untruncated modes M, the LIF-RCWA shows no violation of the law of energy conservation.

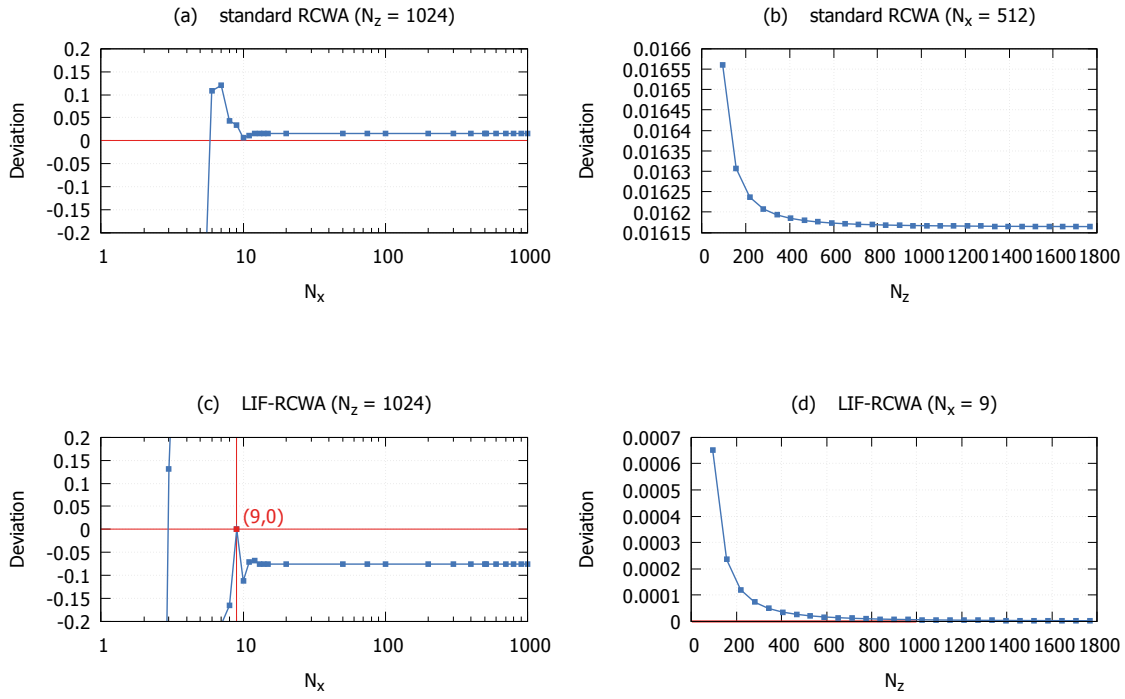


Figure 5.11: Influence of the sampling on the energy conservation behavior of the standard RCWA and the LIF-RCWA. In both cases the permittivity and the fields were represented by $M_0 = 4 \Rightarrow M = 9$ Fourier modes.

5.3.2 Energy conservation in lossless gratings

The question remains why energy conservation seems always be ensured for a lossless gratings even if an inconsistent truncation scheme is used in the standard RCWA. In fact, the value of $1 - \mathcal{R} - \mathcal{T}$ strongly depends on the properties of the convolution matrix $[\![\epsilon]\!]$ in Eq. 3.8. The following listing summarizes some intuitive and empirical facts:

1. $\epsilon \in \mathbb{R}$ (non-absorbing grating) $\wedge$ $[\![\epsilon]\!]$ is Toeplitz $\Rightarrow$ $[\![\epsilon]\!]$ is hermitian $\Rightarrow 1 - \mathcal{R} - \mathcal{T} = 0$
2. $\epsilon \in \mathbb{R} \wedge$ $[\![\epsilon]\!]$ is *not* Toeplitz (*no physical grating*) $\wedge$ $[\![\epsilon]\!]$ is hermitian $\Rightarrow 1 - \mathcal{R} - \mathcal{T} = 0$!
3. $\epsilon \in \mathbb{C}$ (absorbing grating) $\Rightarrow$ $[\![\epsilon]\!]$ is Toeplitz $\wedge$ $[\![\epsilon]\!]$ is not hermitian $\Rightarrow 1 - \mathcal{R} - \mathcal{T} \neq 0$
   - 3.1. $[\![\epsilon]\!]$ is cyclic $\Rightarrow 1 - \mathcal{R} - \mathcal{T} =$ integrated absorption
   - 3.2. $[\![\epsilon]\!]$ is not cyclic $\Rightarrow 1 - \mathcal{R} - \mathcal{T} \neq$ integrated absorption

Statement 2 is backed by empirical evidence: It was found that in at least some, if not all, cases it is sufficient if $[\![\epsilon]\!]$ is an arbitrary hermitian matrix and the condition for energy conservation $1 - \mathcal{R} - \mathcal{T} = 0$ will be met as a consequence - even if $[\![\epsilon]\!]$ is not a Toeplitz matrix and thus does not correspond to a meaningful grating. This is similar to the case of inconsistent truncation of $[\![\epsilon]\!]$ in the standard RCWA, which also always leads to $1-\mathcal{R}-\mathcal{T} = 0$. Hence, it can be concluded that the verification of $1-\mathcal{R}-\mathcal{T} = 0$ is not a sufficient condition to decide, whether the calculation is correct or not. However, statement 3.1 and 3.2 indicate that the cyclic approach provides the correct truncation scheme as the energy deficit can fully be explained by absorption in a lossy medium.

5.3.3 Conclusion

The above studies show that there is a discrepancy in the standard RCWA between the net power loss at an absorbing grating and the amount of power that can be addressed to the process of absorption. This is a violation of the law of energy conservation and evidence indicate that the reason is the inconsistent mode truncation in the Toeplitz matrices of the grating's permittivity (and permeability). If in contrast the Toeplitz matrices are defined in a cyclic manner as introduced in Chap. 3 the deviation between the two quantities only dependent on the sampling of the electromagnetic fields and converge towards zero if the sampling is increased.
Furthermore, even for non-absorbing gratings the findings indicate that the use of non-cyclic Toeplitz matrices may cause some inaccuracies that usually go undetected, since the criterion for energy conservation $1 - \mathcal{R} - \mathcal{T} = 0$ does not ensure a correct calculation.

Chapter 6

Design concepts and realization

6.1 Motivation and idea

The integration of optical components in standard semiconductor technologies is still a challenge of today's chip industry. Although optics can offer superior speed and data throughput without the problems of conventional electronics like crosstalk and high damping rates in wires, and although some professional applications like telecommunication backbones and high performance networks already benefit from the various advantages of optical data transmission, up to the present day the industry still struggles in applying optical components to consumer products and chip level applications[118]. A successful integration would allow cheap mass production and advanced applications like optical chip-to-chip communication or optical clock distribution networks. One of the biggest obstacles, though, is an impending change of technology. It is feared that the optical integration might require large investments by the industry, since new technologies and processes might be required, while established equipment and knowhow could become obsolete. These days a similar development can be observed in the challenging transition from UV-lithography towards EUV-lithography, which requires a new generation of entirely redesigned exposure units[9].

In very-large-scale integration (VLSI) chips the complementary metal-oxide-semiconductor (CMOS) technology is by far the most common one[33]. By using complementary pairs of positive- and negative channel field effect transistors for logic functions (pFETs and nFETs), from which one is always turned off, the significant amount of power is only drawn during switching states. This saves a lot of energy compared to other technologies and also reduces heat generation and noise dependence[34]. For these reasons, it would be most desirable to be able to integrate optical devices like emitters, detectors, optical modulators and also passive structures like waveguides into CMOS technology. The focus of this chapter is the development of a compatible photodetector.
A special extension of the CMOS technology is called silicon-on-insulator (SOI), which uses an insulating layer (e.g. silicon dioxide) inside the silicon substrate in order to reduce parasitic capacitance. It thereby lowers device switching times, power consumption as well as leakage currents. On the optical side, the crystalline silicon layer, which lies on top of the insulator can be used for optical waveguiding or other passive devices for integrated optics[95]. Both facts make SOI an interesting support technology for this approach.

However, active optical components primarily rely on materials or compounds with high absorption or emission coefficients, which are usually those that have a direct band gap between the valence band and the conduction band. Typical representatives are III-V semiconductors

like Boron-, Aluminium-, Gallium- and Indium-based nitrides, phosphides or arsenides. At a direct band gap a photon with an energy larger than the band gap energy can either be emitted or absorbed directly. Silicon, on the other hand, being the standard material for semiconductor substrates of almost all chip designs today, has an indirect band gap[1]. This means that an extra crystal momentum, which manifests as a quantized quasiparticle called a phonon, is required in order for the absorption or emission process of a photon to happen[96]. This involves the absorption of a phonon in the emission process of a photon. And since the presence of an additional phonon is less likely to occur, the emission efficiency of silicon is significantly decreased compared to a direct semiconductor.
It is also not straighforward to integrate direct III-V semiconductors on silicon substrates. Since the band gap energy of a semiconductor is directly linked to its lattice constant[24] and since the lattice constants of III-V semiconductors differ from that of silicon, their direct contact can cause piezoelectrical charges, dangling bonds and actual damage in the material. Therefore, III-V semiconductors are not part of a standard CMOS process and require additional post-processing steps[2].
The absorption process of a photon in silicon, on the other hand, does not depend on the accidental appearance of a phonon. Instead the conservation of momentum, which holds in electrodynamics as well as in quantum mechanics, always predicts the *emission* of a phonon when an electron from the conduction band recombines with a hole in the valence band[96]. As a consequence, the photon absorption efficiency of silicon is not decreased and is even superior to that of some direct semiconductors (cf. Fig. 6.1).
In order to generate a measurable photocurrent, the illuminated silicon area must also provide the depletion region of a p-n junction with an electric field separating the charge particles. Luckily, p-n junctions can be found in plenty in CMOS technology and thus silicon CMOS technology can - in principle - be used to build an effective photodetector.

However, CMOS design only provides layer thicknesses on the scale of nanometers, which almost prevents any absorption. Therefore, the basic idea of this chapter is to utilize existing structures of the CMOS-technology as diffraction gratings in order to locally increase absorption and thereby develop a resonant photodetector concept as well as a first prototype that is fully compatible with the standard SOI CMOS process without the need of any post-processing steps. The chapter starts with a preliminary theoretical concept and discusses the challenges and potentials of the project. In a second part a first prototype is presented, which was developed in collaboration with IBM, followed by some experimental results. At the end of the chapter, a low-cost monochromatic light source with variable wavelength is shown, which is considered as an in-house alternative to the expensive tunable laser for photodetector characterization.

[1] Silicon has in fact also a direct bandgap at $4.10\,eV$. However, this corresponds to a wavelength of $0.30\,\mu m$ in the ultraviolet, which has low transmittance in common fiber optics.
[2] Some experimental processes actually do exist as published recently by IBM[111]

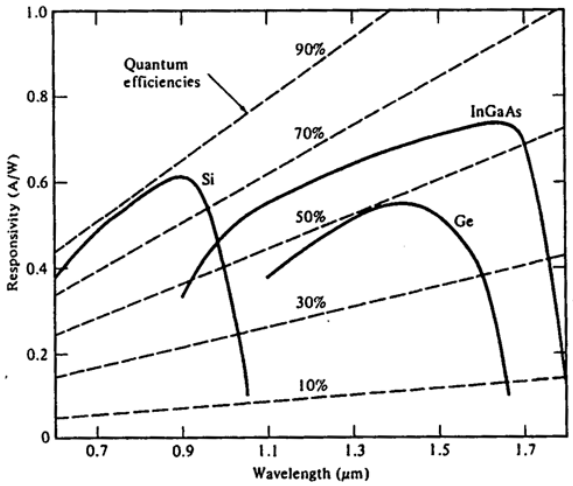


Figure 6.1: Responsivity curves for Si, Ge, and InGaAs. Source: Agrawal, G.P., *Fiber-Optic Communication Systems*, John Wiley & Sons, New York, (1997)

6.2 Concept for a CMOS compatible photodetector

6.2.1 First considerations, challenges and potentials

An effective photodetector primarily needs to provide a strong electric response to the incident light and thus requires a high photon-electron conversion rate. This rate is described by the external quantum efficiency η_e as the average ratio between the number of electrons N_e, which contribute to the photocurrent, to the number of photons N_{ph}, which illuminate the photodetector at a specific energy - at a steady flow:

$$\frac{N_e}{t} = \eta_e \frac{N_{ph}}{t} \tag{6.1}$$

Taking into account only the absorbed photons leads to the internal quantum efficiency η_i:

$$\frac{N_e}{t} = \eta_i \frac{N_{abs.ph}}{t} = \eta_i \eta_{abs} \frac{N_{ph}}{t} \tag{6.2a}$$

$$\Rightarrow \eta_e = \eta_{abs} \eta_i, \tag{6.2b}$$

where η_{abs} is the absorption efficiency P_{abs}/P_{in} (cf. Eq. 5.20) inside the illuminated region of the photodetector.

Considering the flow of electric particles (photocurrent I) and photon energy (radient flux P_{in}),

$$I = \frac{N_e q}{t} \tag{6.3a}$$

$$P_{in} = \frac{N_{ph}\hbar\nu}{t} \tag{6.3b}$$

where q is the electric charge of an electron, $\hbar$ is the Planck constant and ν is the frequency of the incident light, the photocurrent of a detector can be described as

$$I = P_{in}R \quad \text{with} \quad R = \eta_{abs}\eta_i\lambda\frac{q}{\hbar c} \tag{6.4a}$$

$$\Leftrightarrow \quad I = P_{abs}\eta_i\lambda\frac{q}{\hbar c} \tag{6.4b}$$

with R known as the responsivity of the photodetector. Hence, a strong signal depends on the optical incident power P_{in}, the wavelength λ, the internal quantum efficiency η_i of the photodetector material as well as its absorption efficiency η_{abs}.

In a working application the optical incident power is usually supposed to be small in order to avoid heating problems and to reduce overall energy consumption. Since, silicon has an indirect band gap of $1.12\,eV$[121], which means that photons with higher energy or with a wavelength below the long wavelength cutoff of about $1100\,nm$ are actually able to excite electrons across the material's band gap. Due to absorption bands in optical fibers, typical wavelengths for telecommunication systems and other long distance network connections are $850\,nm$, $1310\,nm$ and $1550\,nm$[8]. The corresponding light sources are therefore easily available and a reasonable choice for a silicon detector would be $850\,nm$. Finally, the internal quantum efficiency of the detector is usually below one, since even the energy of an absorbed photon can convert into a phonon or reradiate thermally. However, silicon has a long minority carrier lifetime in the range of milliseconds compared to other semiconductors like Gallium Arsenide (GaAs) of only few nanoseconds[121]. This increases the probability of an electron to account for an effective photocurrent and also explains the superior responsivity of a silicon based detector in Fig. 6.1.

Besides that, the internal quantum efficiency also depends on the distribution of incident photons in the photosensitive area and the diffusion length of the free charge carriers: On either side of the junction of a positively (p-) doped and a negatively (n-) doped silicon semiconductor a depletion region forms due to the recombination of excessive free electrons and holes (minority charge carriers). Inside this depletion region, the remaining atoms of the crystalline solid (majority charge carriers) create an electric field, which separates the newly generated electron hole pairs by pulling any free electron to the n-site and any hole to the p-site (also known as drift) contributing to a displacement current in the external circuit. Since the electric field primarily extends over the depletion region and charge carriers in

the outside are only moved by the slow diffusion process, their probability of spontaneous recombination is considerably increased, lowering the overall internal quantum efficiency. Thus, one way of increasing the efficiency is to increase the relative size of the depletion region in comparison to the overall extend of the photodetector. This is the concept of a pin-diode, which hosts an additional weakly or undoped region (also called intrinsic region) with no charge carriers between its two highly doped p- and n-regions. It thereby also increases the detector's response time, since the charge carriers are moving faster in the electric field than in the outside. Moreover, the increased distance between p- and n-region leads to a lower capacitance ($C = \epsilon\epsilon_0 A/W$, with A as the junction profile and W as the width of the depletion region) and therefore improves isolation and minimizes reverse leakage current.
So finally, the absorption efficiency is the only parameter left to effectively improve the performance of the considered concept. Usually, however, absorption is only determined by the detector's raw material, the wavelength of the incident light and the penetration depth. But as mentioned before, this only leads to very weak absorption in the thin CMOS silicon layers. To put it in numbers: At a wavelength of 850 nm only 2 % of the incident light would be absorbed in a 80 nm thin layer of polycrystalline silicon as visualized in Fig. 6.2. An increase in the thickness of the active layers is also not an option, since most design parameters including very small layer heights are fixed by constraints of the considered 45 nm CMOS technology.

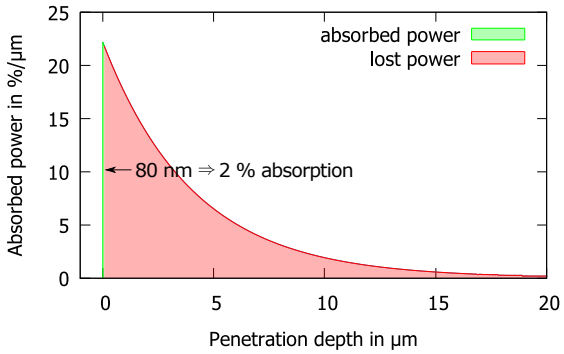


Figure 6.2: Illustration of the absorbed power in a polysilicon layer against the penetration depth of the incident photons

6.2.2 Design and simulations

Fig. 6.3 shows a simplified SOI CMOS layout provided by IBM, the partner of this project. It illustrates from the bottom up a layer of polysilicon as the substrate with a permittivity of $\epsilon = 13.81$, followed by a 145 nm thin layer of silicon oxide (called buried oxide or BOX) with $\epsilon = 2.11$ to electrically isolate against parasitic leakage currents, an 80 nm thin polysilicon layer with $\epsilon = 13.81 + 0.145i$, which can be variously doped and potentially be used as a photoactive layer, and an 85 nm thin grating structure, which is formed by the polysilicon transistor gates of the MOSFETs of CMOS technology. Finally, the entire layout is covered by silicon nitride with $\epsilon = 3.971$.

These small dimensions combined with the weak absorption property of silicon seem to contradict the goal of maximizing the overall absorption in the photoactive layer. In fact, the formula for local absorption (cf. Eq. 5.18, 6.5) suggests to aim for high absorption property $\mathrm{Im}\left(\epsilon\left(\mathbf{r}\right)\right)$ and large volume V:

$$P_{\mathrm{abs}}\left(\mathbf{r}\right) = \frac{k_0}{2Z_0}|E_0|^2 \int_V \mathrm{Im}\left(\epsilon\left(\mathbf{r}\right)\right)|\mathbf{E_1}(\mathbf{r})|^2 \, d^3V \tag{6.5}$$

However, the introduction of the diffractive structure (cf. Fig. 6.3) can significantly improve the relative intensity $|\mathbf{E_1}(\mathbf{r})|^2$ in the junction! Using the transistor gates as a diffraction grating on top of the photodiodes the incident light can be concentrated and thereby (relatively) enhanced in the depletion region resulting in more photons being absorbed and a higher photocurrent being generated (cf. Eq. 6.4b). Therefore, the goal is to determine those values of the additional design parameters (gate width (w) and grating period (P), cf. Fig. 6.3), which lead to maximal light absorption in the active channel and preferably in the depletion region.

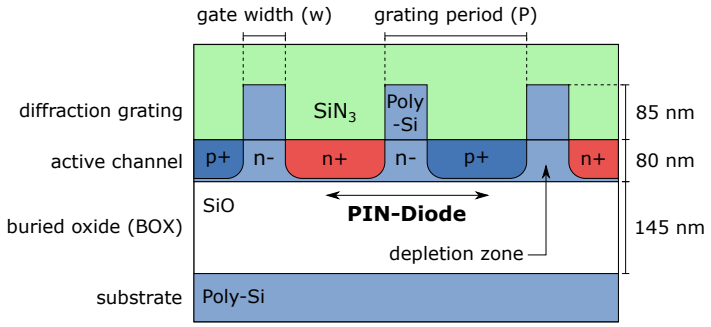
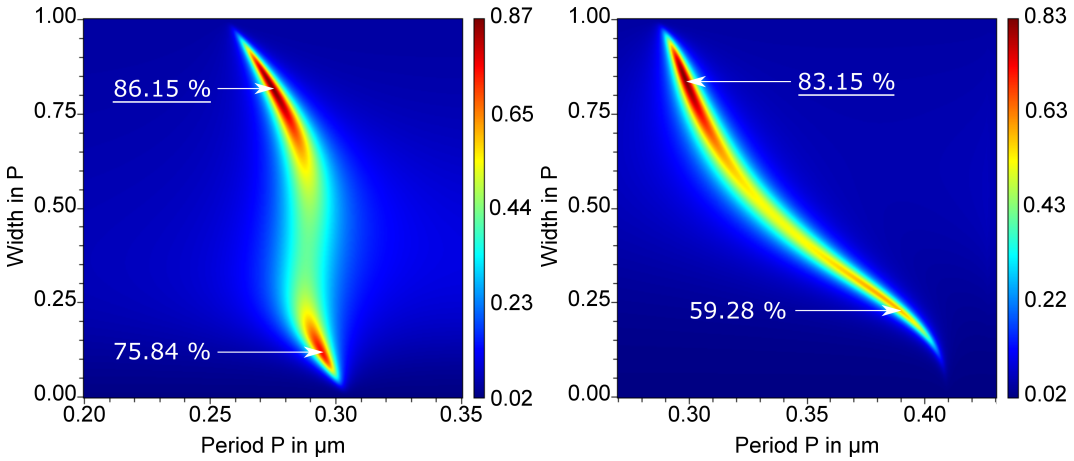


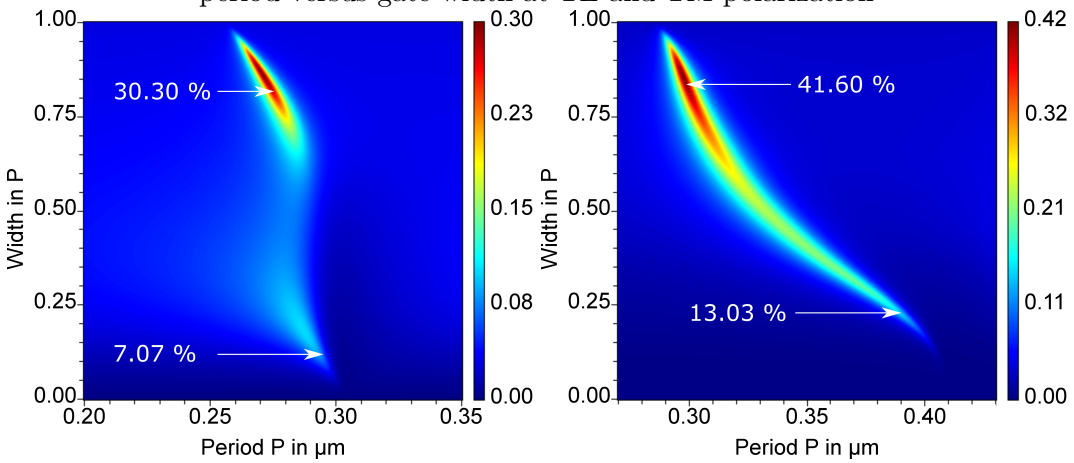
Figure 6.3: Layout of a CMOS compatible photodiode with transistor gates utilized as a diffractive grating structure

The first simulation results shown in Figs. 6.4a and 6.4b illustrate the global absorption in the layer stack for TE and TM polarization in two-dimensional parameter scans over grating period and gate width.
The results are indicating two things: For one, the diffraction grating is capable of increasing the total absorption from 2 % up to 86 % for TE polarization and from 2 % to 83 % for TM polarization. However, it is not clear from this simulation how much light is actually being absorbed in the active layer.
On the other hand, the absorption maxima for TE and TM polarization do not match and high gains for either polarization type even exclude one another. This is unfortunate, since it divides the absorption efficiency for unpolarized incident light in half. Nevertheless, application designs can be specified for one polarization type only.

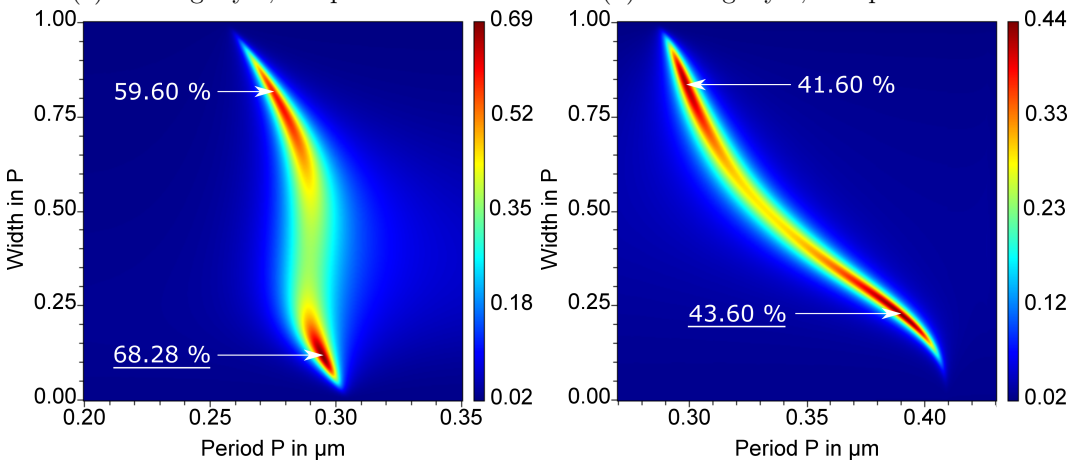


(a) TE polarization (b) TM polarization

Figure 6.4: Global absorption depicted for two-dimensional parameter scans of grating period versus gate width at TE and TM polarization

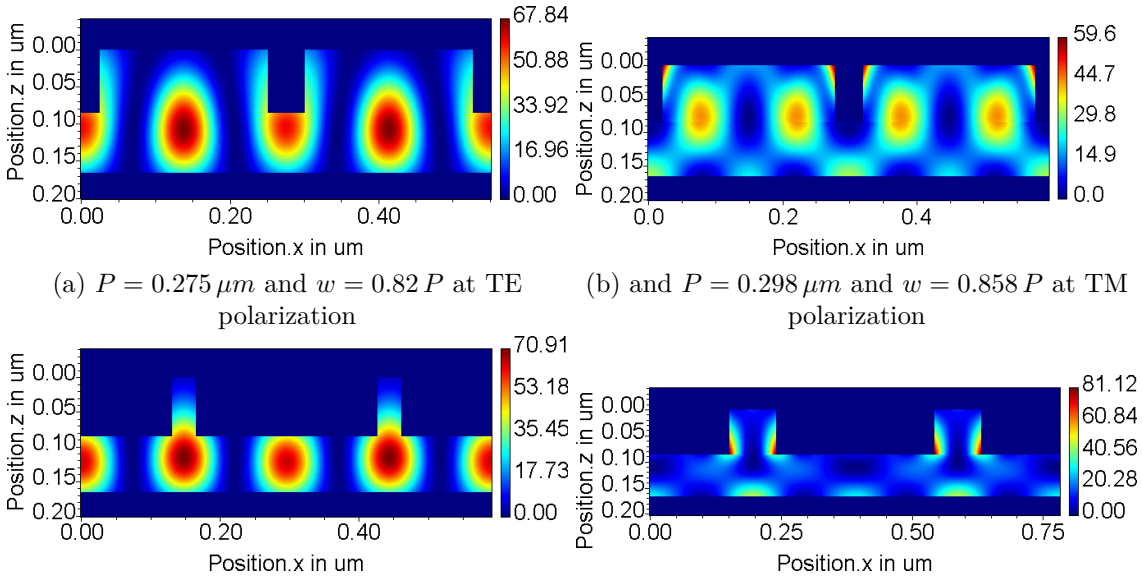


(a) Grating layer, TE polarization (b) Grating layer, TM polarization



(c) Active layer, TE polarization (d) Active layer, TM polarization

Figure 6.5: Layerwise accumulated local absorption in a two-dimensional parameter scan of grating period versus gate width for TE and TM polarization



(a) $P = 0.275\,\mu m$ and $w = 0.82\,P$ at TE polarization

(b) and $P = 0.298\,\mu m$ and $w = 0.858\,P$ at TM polarization

(c) $P = 0.295\,\mu m$ and $w = 0.113\,P$ at TE polarization

(d) $P = 0.390\,\mu m$ and $w = 0.225\,P$ at TM polarization

Figure 6.6: Simulation of the local absorption in the grating structure and the photoactive layer

Figs. 6.5a-6.5d show the same parameter scans as before, but this time for the accumulated local absorption in each of the first two absorbing layers. The simulations reveal, that those parameter combinations, which show the highest values of global absorption (at $P = 0.275\,\mu m$ and $w = 0.82\,P$ for TE polarization and $P = 0.298\,\mu m$ and $w = 0.858\,P$ for TM polarization) do in fact not match with the parameters for maximal absorption in the active layer ($P = 0.295\,\mu m$ and $w = 0.113\,P$ for TE polarization and $P = 0.390\,\mu m$ and $w = 0.225\,P$ for TM polarization) - for neither type of polarization.

A closer look at the energy distribution inside the detector is shown in Figs. 6.6a - 6.6d for all discussed parameter configurations with absorption maxima. In either case, the results, which have been optimized for TE polarized light, seem to be better suited for a photodetecting application, since the light absorption concentrates right under the gates in the middle of the depletion zone, maximizing the charge carrier generation.

6.2.3 Summary

The here presented layout was actually fabricated using IBM's 45 nm SOI CMOS technology by Moll et al.[93] in advance of this work. Only the rigorous simulations of the layerwise local absorption were not available at the time, as the appropriate tools were first developed during this work and later published in [5].

The results are again summarized in Tabs. 6.1 and 6.2. They show that a simple comparison of global absorption values can in fact be misleading and the efficiency of a photodetector design can be further improved by calculating the local absorption in the photoactive region. Nevertheless, in this scenario the maximum achieved absorption values in the active channel do not differ significantly.

	without grating	with grating	
		TE polarization	TM polarization
grating absorption	0 %	30.3 %	41.6 %
channel absorption	2 %	59.6 %	41.6 %

Table 6.1: Overview over the absorption values in the grating and the active channel, if only *global absorption* is optimized

	without grating	with grating	
		TE polarization	TM polarization
grating absorption	0 %	7.07 %	13.03 %
channel absorption	2 %	62.28 %	43.6 %

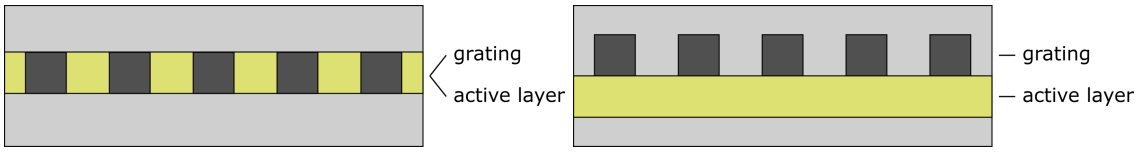
Table 6.2: Overview over the absorption values in the grating and the active channel, if *local absorption* is optimized

The study shows that the precise knowledge about the local absorption rather than a global absorption can be very useful. Detailed insight in the absorption process enables the development of better optical components. In this case a grating structure on top of the photodiode was optimized for maximal absorption in the depletion region.
Another conclusion of this analysis is the difficulty to use a one dimensional grating structure for the development of an enhanced photodetector that performs comparably well for both types of polarization (TE and TM). Nevertheless, it was shown that intrinsic features of the CMOS technology can be utilized to significantly improve the performance of a fully CMOS compatible photodetector.

6.3 Realization of a CMOS compatible photodetector

6.3.1 Motivaton and idea

Due to a design change in IBM's SOI CMOS technology, which involved the replacement of the transistor gate material from polysilicon (Poly-Si) to titanium nitride (TiN), the former approach (cf. Sec. 6.2) could not be pursued any further. Since the new metal gates would be very strong absorbers itself, they would significantly decrease the photodetector's performance when put on top of the active layer.
Hence, the new approach utilizes the shallow trench isolation (STI)[3] as an *in-place* resonance grating inside the photoactive layer (cf. Figs. 6.7a) and replaces the role of the transistor gates, which served as an *on-top* diffraction grating in the previous design (cf. Fig. 6.7b). A secondary bulk diode can be used for maximal photon collection, increased overall charge carrier generation and high signal strength.



(a) *in-place* resonant grating (b) *on-top* grating on a wave guide resonator

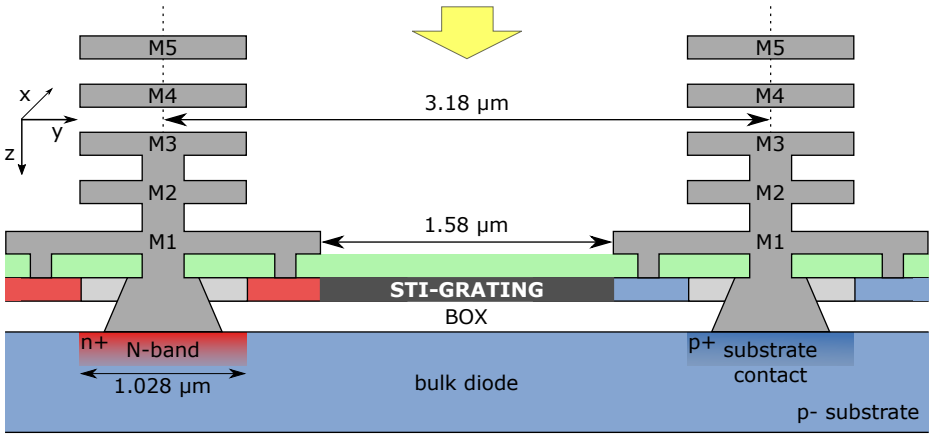
Figure 6.7: Two types of grating structures

This project was planned and carried out in cooperation with M. Fertig from IBM Research & Development Böblingen and N. Moll, T. Morf, T. Stöferle and J. Hofrichter from IBM Research Laboratory Zurich. While IBM provided the manufacturing technology and the expertise in chip design, all simulations regarding the optical properties of the photodiode, such as the design of the resonance grating and the calculation of local absorption, were part of the present research work.

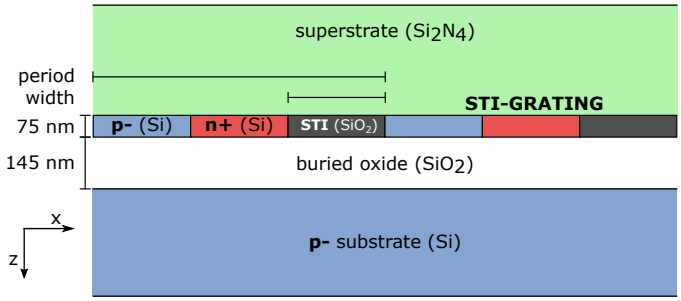
6.3.2 Design

Fig. 6.8a shows a cross section of the full layout. M5 to M1 indicate the different metalization layers, which are used for routing and for connecting the vertical contacts (at a distance from the photoactive aperture). Under a cover layer of silicon nitrogen (Si_3N_4) lies the STI grating, which is formed by the shallow trench isolations (STI) inside a polysilicon (Poly-Si) layer. The shallow trenches are made of silicon dioxide (SiO_2) and are usually used to electrically isolate the transistors against each other. The polysilicon serves as the photoactive material and can be doped variously within IBM's SOI technology constraints in order to create multiple p-n junctions (cf. Fig. 6.8b). The layer's height, however, is limited to only 75 nm

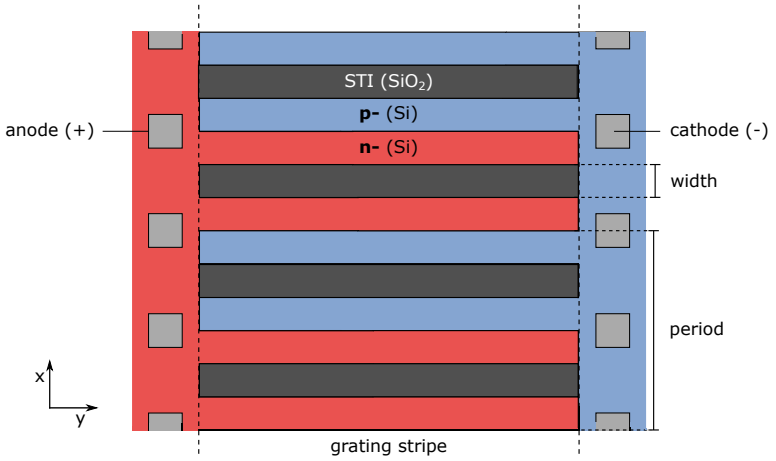
[3]Shallow trench isolation is also part of IBM's SOI CMOS technology.



(a) Cross section of the photodetector layout



(b) Cross section of the STI grating



(c) Top view of one stripe of the STI grating with contacting

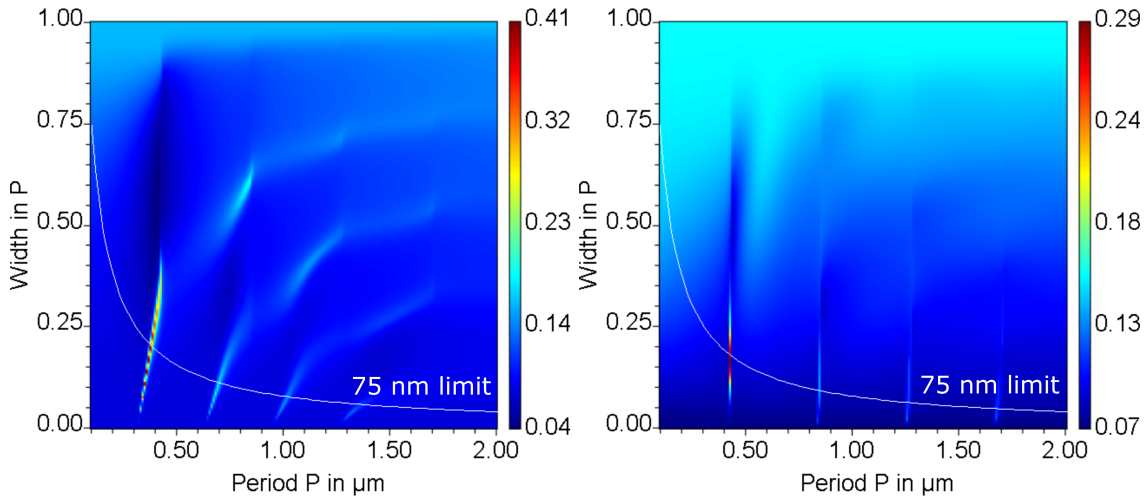
Figure 6.8: Different views of the photodetector layout

due to the underlying 145 nm thick oxide (BOX), which isolates the transistors against a

joint substrate to avoid latch-ups. This fact, together with another technology constraint that requires the silicon dioxide trenches to be larger than 75 nm, minimizes the available volume in the p-n regions for the purpose of light absorption and emphasize the necessity for a maximum light-sensitive depletion volume. The slightly p-doped polysilicon substrate and the highly n-doped n-channel FETs form another layer of buried bulk diodes (cf. Fig. 6.8a). Due to a maximal distance of 3.18 μm between the vertical contacts, repeating stripes of 1.58 μm in width are available for the STI grating. One stripe is shown in Fig. 6.8c including its electric contacts. Anodes and cathodes are placed in an alternating pattern between the stripes, so that electrodes and metal wiring is shared by neighboring diodes in order to maximize the active region of the detector. The implant of the bulk diode contact has a width of 1.028 μm leading to a depletion region of approximately 1 μm.

6.3.3 Preliminary simulations

The photodetector design is again intended to work for a wavelength of 850 nm, which is a standard in optical communication (cf. Sec 6.1). Hence, simulations were conducted assuming the following optical material properties: $\epsilon = 4.08$ for the non-absorbing silicon nitride superstrate, $\epsilon = 13.36 + 0.025i$ for the photoactive silicon, $\epsilon = 2.1$ for the non-absorbing silicon, which form the shallow trenches as well as the buried oxide and $\epsilon = 13.41 + 0.07i$ for the absorbing p-doped silicon substrate.



(a) TE polarization (b) TM polarization

Figure 6.9: Global absorption depicted for two-dimensional parameter scans of grating period versus oxide section width at TE and TM polarization. A minimal oxide section width due to CMOS technology constraint of 75 nm is indicated in white.

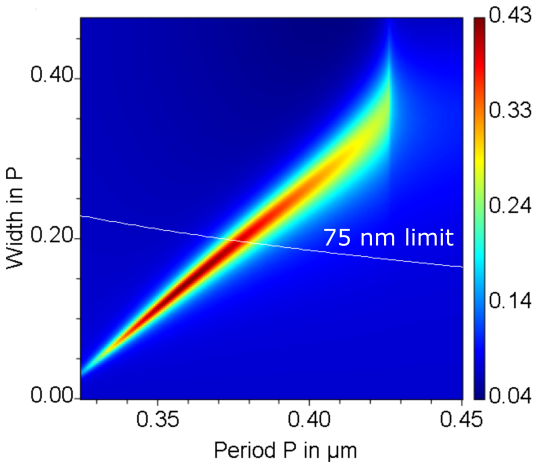


Figure 6.10: Close-up of the two-dimensional parameter scan of grating period versus oxide section width (at TE polarization)

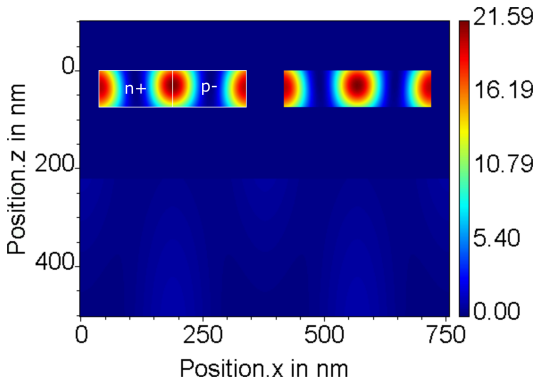
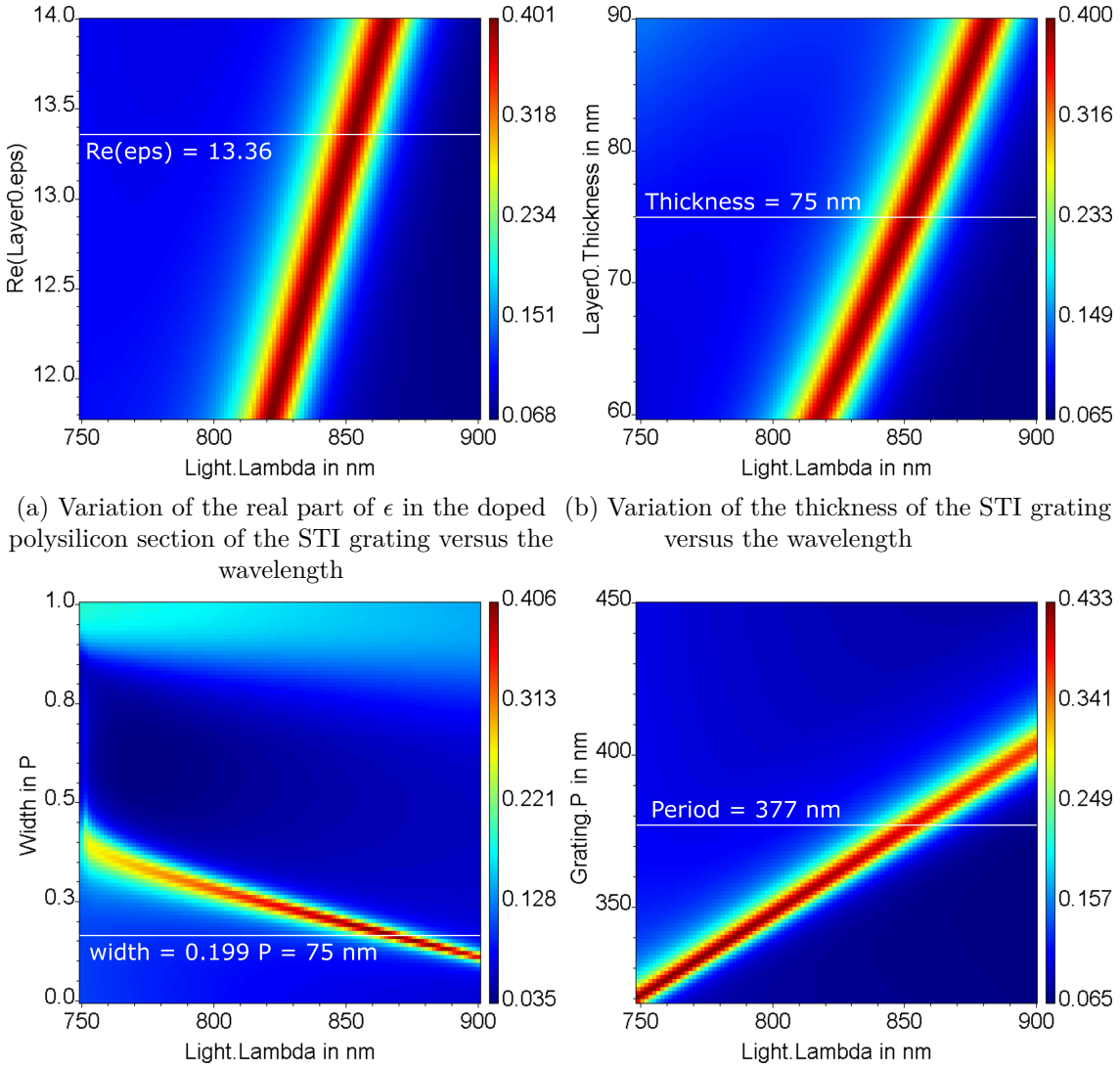


Figure 6.11: x/z-cross section of the local absorption at optimal parameter configuration ($P = 377\,nm$, $w = 0.199\,P = 75\,nm$) (at TE polarization)

Figs. 6.9a and 6.9b show the calculated global absorption in the design for a variation of the two crucial parameters: the period P of the STI grating and the relative width of the silicon oxide sections w. The optimal parameters are determined by the technology constraint for a minimal oxide section width of $75\,nm$. In case of TE polarized light this leads to a grating period of $P = 377\,nm$ and an oxide section width of $w = 0.199\,P = 75\,nm$ (Fig. 6.10 shows a closer view). In case of TM polarized light the optimal values are $P = 425\,nm$ and $w = 0.176\,P = 75\,nm$. In these particular configurations the corresponding absorption rates are 39.5 % (for TE polarization) and 28.5 % (for TM polarization). Since, again, no parameter configuration provides high absorption values for both types of polarization, the further investigation concentrated on pure TE polarization, which reaches higher overall absorption values.

Fig. 6.11 shows the distribution of absorbed power in the photodetector with a high concentration in the p-n junction, which is desirable for a high responsivity. The accumulated absorption in the resonant STI grating and the bulk diode are 23.92 % and 15.61 %, respectively. However, since all the absorbing structures in the design are either part of the top diode or the bulk diode, the present approach aims for high overall absorption. Hence, there is no purpose for another parameter scan to optimize one of the diodes separately, because it would always decrease the efficiency of the other one.

Nevertheless, further simulations were performed in order to estimate the influence of different parameter deviations, which might occur during the fabrication process. Figs. 6.12a and 6.12b illustrate that a variation of the thickness or permittivity of the grating slightly shifts the wavelength dependency away from $850\,nm$, but does not affect the maximum absorption. A variation of the grating period and especially of the duty cycle on the other hand, causes a considerably stronger wavelength shift and also influences the efficiency of the detector (Figs. 6.12c and 6.12d).



(a) Variation of the real part of ϵ in the doped polysilicon section of the STI grating versus the wavelength

(b) Variation of the thickness of the STI grating versus the wavelength

(c) Variation of the relative width of the oxide sections in the STI grating versus the wavelength

(d) Variation of the period of the STI grating versus the wavelength

Figure 6.12: Impact of different design parameter variations on the detector's wavelength dependency

6.3.4 Fabrication

The TE-design of the photodetector (cf. Fig. 6.11) was manufactured as a unit of $25 \times 25\,\mu m^2$ in size on a silicon wafer using a standard $32\,nm$ STI CMOS process, which involved activation annealing and an advanced immersion lithography for precise channel length control. Doping levels were approximately 10^7 in the p^- doped regions and 10^{19} in the n^+ doped regions. Fig. 6.13 shows the side view of the manufactured STI grating taken by a Scanning Electron Microscope (SEM). It reveals multiple parameter deviations from the design.

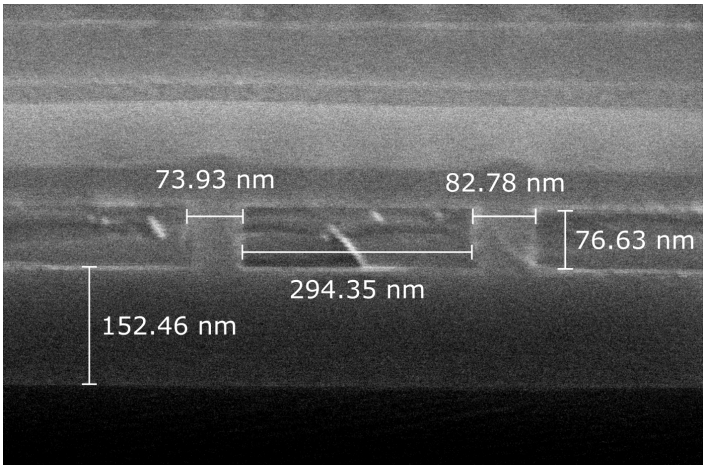


Figure 6.13: SEM image of the STI grating (side view) with measured dimensions

In order to simulate the effects of these deviations, it was assumed that the average grating period was realized as $377\,nm$ according to the design specification. Since the grating period depends on the very precise lithographic process rather than on the development chemistry this assumption is reasonable. The average width of the oxide sections was therefore estimated to be $\varnothing w = 0.2078\,P = 78.36\,nm$, which, together with a measured grating thickness of $76.63\,nm$ and a buried oxide height of $152.46\,nm$, results in a shift of the maximal absorption towards smaller wavelengths as shown in Fig. 6.14.

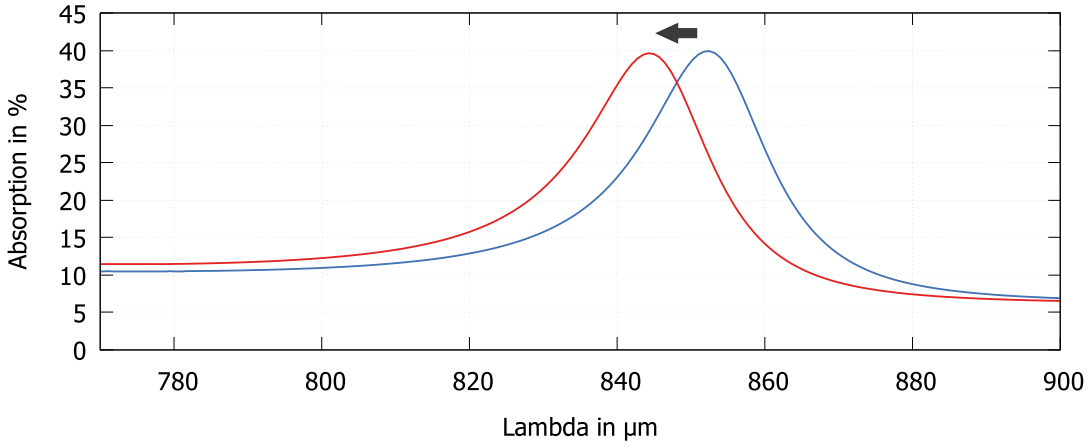


Figure 6.14: Wavelength shift of the photodetector absorption peak due to parameter deviations in the manufacturing process. Absorption for design parameters (blue) versus absorption for measured parameters (red)

6.3.5 Measurement and Results

The photodetector was characterized using a 80-fs-pulse, linearly polarized laser with a repetition rate of 80 MHz. The light source contained an optical parametric oscillator (OPO), which was pumped by a mode-locked Ti:sapphire laser, in order to allow for a wavelength tuning between 770 nm and 900 nm. The emitted light was then guided by an optical fiber and collimated in free space through an attenuator and onto the detector device using a lens with a long focal length (125 mm) to avoid a wide angular spectrum. At the position of the sample the output power was measured to be about 30 μW. In order to receive the electric response of the detector, broadband (DC-50 GHz) ground-signal-ground probes from GGB Industries were used, which contacted the wafers pads.

The short duration of the laser pulse of only 80-fs leads to a spectral broadening. Assuming a Gaussian shaped pulse, the time-bandwidth product $\triangle t^{\mathrm{FWHM}} \triangle f^{\mathrm{FWHM}}$ is known to be greater than 0.441[132]. Thus, a spectrum broadening of at least $\triangle \lambda^{\mathrm{FWHM}} \geq 0.441 \lambda_c^2 / (\triangle t^{\mathrm{FWHM}}\, c)$ $= 13.285\, nm$ can be assumed for a center wavelength of $\lambda_c = 850\, nm$. Fig. 6.15 shows the expected change in wavelength dependency based on the convolution with an intensity distribution of a normalized Gaussian shaped spectrum with a full width at half maximum (FWHM) of 13.285 nm.

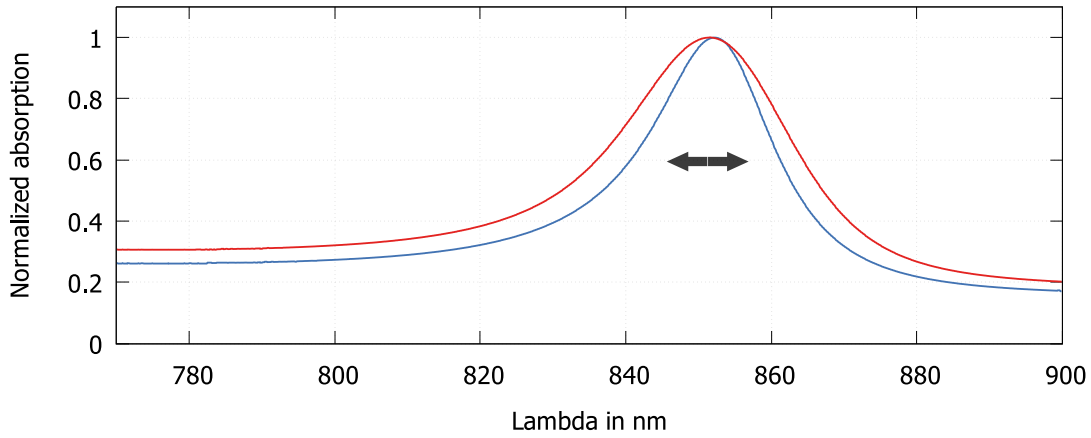


Figure 6.15: Transformation of the photodetector absorption behavior due to the frequency broadening of the pulsed laser source. Absorption at a monochromatic light source (blue) versus absorption at a broadened spectrum (red)

The result of the actual measurement is shown in Fig. 6.16, where the peak responsivity of 0.26 A/W was observed at a wavelength of 830 nm (red line with dots). The blue solid line depicts the expected absorption behavior with the considered broadening effect and a maximum value at 850 nm. The blue dotted line shows the shifted wavelength dependence due to the measured deviations to the design parameters. The remaining displacement to the solid green curve with a maximum value at 830 nm can in fact have various causes as demonstrated in the simulations shown in Figs. 6.12a-6.12d. In this particular case, the match

was achieved with a relative oxide width of $w = 0.23\,P = 86.71\,nm$. This is significantly larger than the SEM measurements of $w = 73.93\,nm$ and $w = 82.78\,nm$. However, the SEM measurements of the two oxide sections already spread by almost 12 %. Another source of uncertainties are the higher layers above the silicon nitride superstrate (cf. Fig. 6.8a), whose compositions were not accessible to this work.
The overall incline in responsivity towards shorter wavelengths aligns with the increasing absorption efficiency and can be primarily ascribed to the bulk diode.

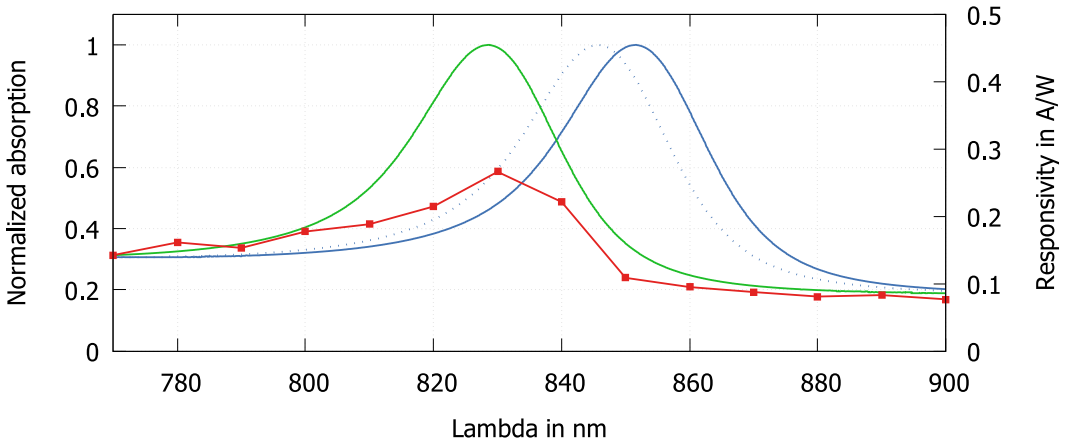


Figure 6.16: Responsivity of the manufactured photodiode (red line-points), calculated absorption with design parameters [$w = 75\,nm$](solid blue), calculated absorption with measured parameters [$w = 78.36\,nm$](dotted blue), calculated absorption with fitted parameters [$w = 86.71\,nm$](solid green)

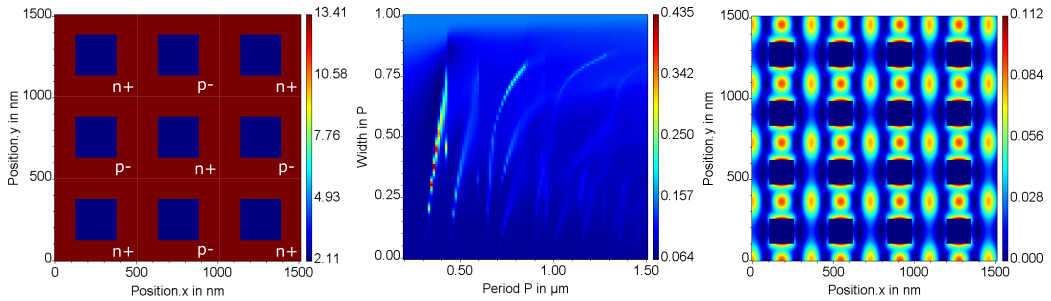
The theoretical maximum for responsivity can be calculated using the previously deduced Eq. 6.4a with $\lambda = 830\,nm$, $\eta_{abs} = 0.395$ and $\eta_i = 1$:

$$R = \eta_{abs}\eta_i\lambda\frac{q}{\hbar c} = 0.264\,A/W$$

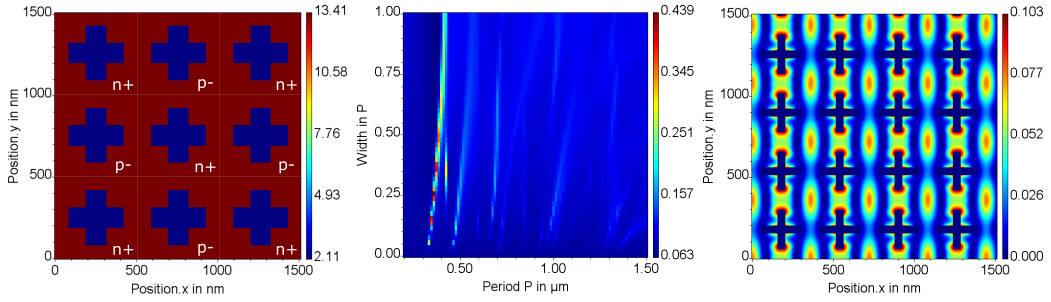
This result shows in fact a very good agreement with the measured responsivity of $0.26\,A/W$ and thus indicates a high internal quantum efficiency η_i as well as an accurate calculation of the absorbed power.

6.3.6 Alternative 2D-grating designs

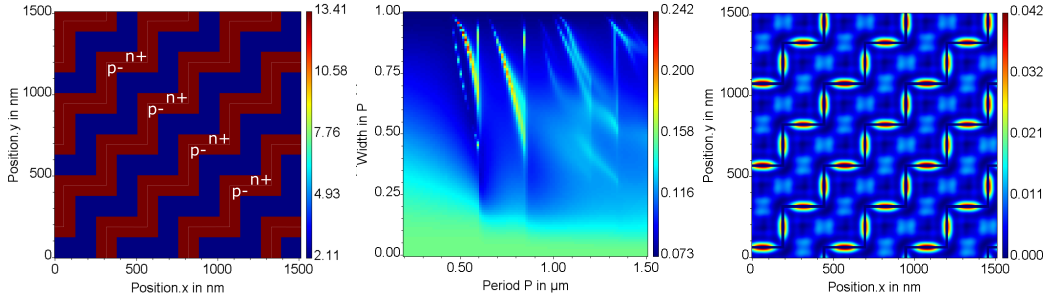
In Sec. 6.3.3 it was shown that a polarization independent design could not be realized with a one-dimensional grating structure. Since light at surface boundaries behaves differently depending on the orientation of the surface as well as the polarization (cf. Sec. 1.1.7), this does not come as a surprise. However, polarization independence can be achieved using a two-dimensional grating design, in which the light encounters the same boundary conditions for both types of polarization (TE and TM).



(a) Grating design (b) Parameter scan (c) Local abs. at $P = 363.8\,nm$, $w = 0.42\,P$ $\Rightarrow$ global abs.: 42.71 %



(d) Grating design (e) Parameter scan (f) Local abs. at $P = 360.3\,nm$, $w = 0.135\,P$ $\Rightarrow$ global abs.: 43.53 %



(g) Grating design (h) Parameter scan (i) Local abs. at $P = 505.4\,nm$, $w = 0.467\,P$ $\Rightarrow$ global abs.: 24.18 %

Figure 6.17: Three examples of polarization independent STI-gratings
first column: 3×3 periods of a grating design
second column: parameter scan of grating period versus oxide section width
third column: top view of distribution of local absorption for an optimal parameter configuration

Figs. 6.17a-6.17i show three examples of two-dimensional designs, which achieve absorption results that are similar to those of the one-dimensional design, which was studied in Sec. 6.3.3 (absorption = 39.5 %). But since the two-dimensional designs are all polarization independent, their actual efficiency could even double in combination with an unpolarized light source.

6.4 Tunable lightsource

6.4.1 Motivation and idea

Without direct access to the high priced, tunable laser source, which was used for the measurements in Sec. 6.3.5, our chair at the University of Heidelberg depended on the stationary equipment of our project colleagues at the IBM Research Laboratory in Zurich. So, in order to achieve the ability of in-house measurements in the long term, we hoped to come up with an affordable alternative light source, which is also tunable over a wide range of wavelengths with an output power in the order of $30\,\mu W$ (similar to illumination power used for the measurements in Sec. 6.3.5) and a small spectral width.
Since an old monochromator was available, an attractive idea was to use a high power broadband light source like a thermal radiator and cut out a narrow window in the spectrum at the desired central wavelength. This approach is investigated in the present section.

6.4.2 Setup

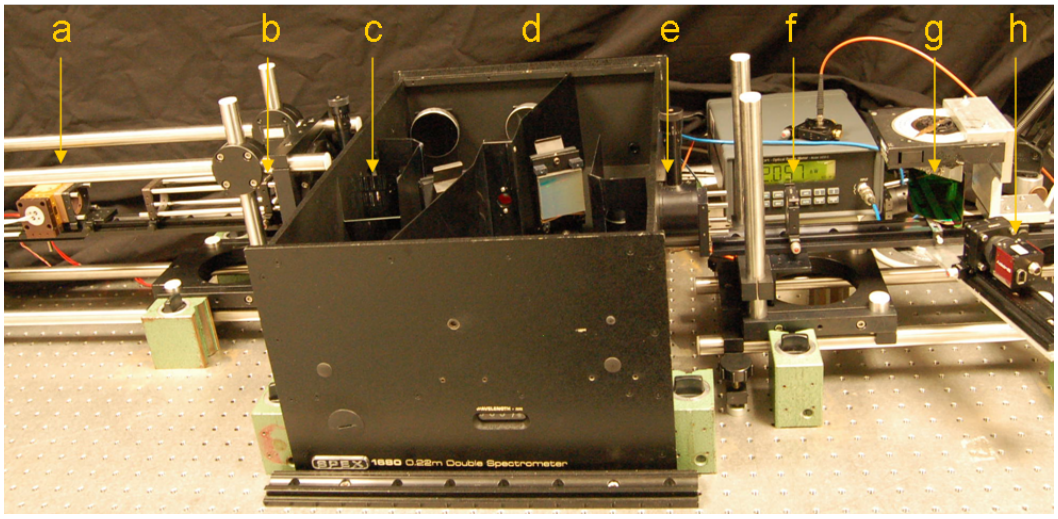


Figure 6.18: Setup with tungsten halogen lamp (a), a first lens (b) to image the source to the entrance slit (c) of a double monochromator (d), and a second lens (f) to image the exit slit (e) again to the sample grating (g). (h) is a calibration camera.

The experimental setup is depicted in Fig. 6.18. A tungsten halogen lamp (a) serves as the broadband light source. It is demagnified and imaged on to the entrance slit (c) of a monochromator (d) using a first achromat (b). The light is then bandpass filtered by the monochromator (SPEX 1680 Double Spectrometer) and leaves the device through the exit slit (e). Here, a second achromat (f) magnifies the image back and thereby reduces the

angular spectrum at the sample. The light finally illuminates a transmissive or reflective diffraction grating (g), which is installed in a rotatable mounting. For calibration purposes, different parts of the setup, like the light source, the grating or its mounting can be replaced by other light sources and measurement devices like

- a helium-neon (HeNe) **gas laser** with a wavelength of 632.816 nm and a spectral width of $\leq 0.5\,nm$,
- an edge emitting **laser diode** with a center wavelength of 636.5 nm and spectral width of 2.2 nm with a collimation lens,
- an **optical power meter** from Newport, Model 1830-C with a sensitivity ranging from 400 nm to 1100 nm (RS-232 serial connection),
  (The wavelength dependency of the internal silicon detector can be corrected semi-automatically by entering the wavelength of the measured input light.)
- a **spectrometer** from mut, Model Tristan USB with a sensitivity ranging from 350 nm to 1100 nm (USB connection),
- a CMOS **camera** from Allied Vision Technologies, Model Guppy F-503 with 2592 × 1944 pixels of 2.2 μm in size and 12 bits for gray shades (firewire connection).
  (The camera can be used to confirm the correct spot position on the sampling grating and also to collect diffracted light from different angles.)

The three sensor devices were plugged into a laptop and fully controlled via custom build software.

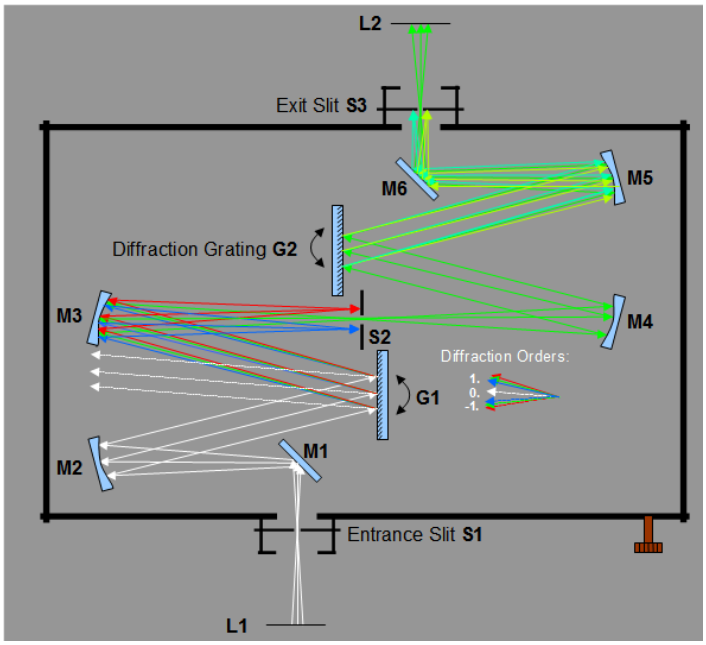


Figure 6.19: Sketch of the inner workings of the SPEX 1680 Double Spectrometer

The particular monochromator as it is shown in Fig. 6.18d and Fig. 6.19 is a so called double monochromator, which works in additive dispersion mode. It basically consists of two series-connected monochromators with coupled diffraction gratings and two slits to increase spectral resolution and to minimize stray light of unwanted wavelengths.
Since the monochromator significantly restricts the spectrum and the radiation angle of the light source, it is important to ensure high efficiency of the system. The limiting factors for the system throughput are the power of the light source, the etendue (also known as geometrical or optical extend, cf. App. A.7), the spectral bandwidth and the general system loss due to a lack of efficiency of the optical components such as lenses, mirrors and gratings.

6.4.3 Calibration

The setup was calibrated with two different lasers: A helium-neon (HeNe) gas laser with a known wavelength of precisely 632.816 nm was used to calibrate position and orientation of the mirrors and gratings inside the monochromator to ensure precise wavelength selection. The laser beam also served as the optical axis, to which all other optical components were aligned. Another edge emitting laser diode with a measured center wavelength of 636.5 nm and spectral width of 2.2 nm was collimated with a lens of 160 mm focal length and the help of a shear plate (cf. Fig. 6.20). Its purpose was to test the focusing and collimation properties of the monochromator and also to determine the instrumental broadening.

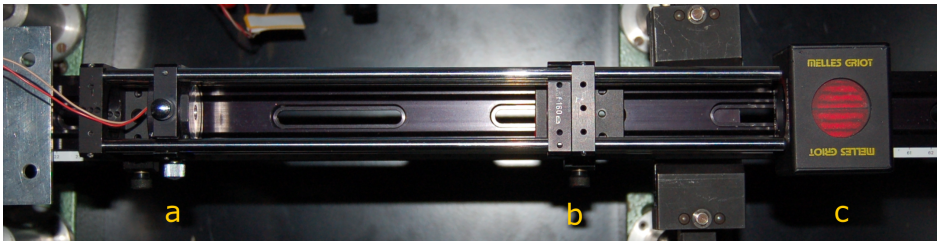


Figure 6.20: Picture of the collimated(b) edge emitting laser diode (a) during the calibration process with a shear plate (c)

6.4.4 Calculation and Dimensioning

The bandwidth and throughput of the setup can be estimated using some simplifying assumptions. Considering the light source to be a Lambertian radiator with an equal radiance in every direction, the system's output power can be defined as

$$P_{out} = L \cdot E, \tag{6.6}$$

where L is the radiance and E is the etendue of the system. The following section aims to derive these quantities and to describe the relevant factors that influence the overall system performance. Finally, the calculated output power is compared with an actual measurement result.

Radiance

The radiance L (in $Wmm^{-2}sr^{-1}$) is defined as the output power per solid angle and per projected area. For thermal radiators the *spectral radiance*, which is defined as the radiance per wavelength, can be obtained by Planck's law via the body temperature T of the radiator. For the tungsten halogen lamp the color temperature was specified to be $T = 3300\,K$. This value can be verified by evaluating the following three equations for the three unknown properties of the light source: the total radiant power P_{ph}, the effective surface area A_{eff} and the color temperature T.

$$P_{ph} = T^4 \sigma A_{eff} \quad \text{with} \quad \sigma = 5.67 \cdot 10^{-8} \frac{W}{m^2} \tag{6.7a}$$

$$P_{ph} = 2\pi A_{eff} L_{max}\,(T) \tag{6.7b}$$

$$P_{ph} = \epsilon P_{el}, \tag{6.7c}$$

Eq. 6.7a describes the Stefan-Boltzmann's law. Eq. 6.7b describes the total radiant power over the full hemisphere and the entire spectrum according to Planck's law (cf. Eq. 6.10). And Eq. 6.7c connects the electric power and the radiant power by assuming the light source to be a gray radiator with a typical emissivity of $\epsilon = 0.425$ for a tungsten halogen lamp[136]. P_{el} is the actual electric power consumption of the lamp and was measured with a multimeter. It led to a voltage of $U = 12.05\,V$, a current of $I = 7.55\,A$ and finally a power consumption of $P_{el} = U \cdot I = 90.978\,W$.

The three Eqs. 6.7a-6.7c are solved by the following values:

$$P_{ph} = 38.665\,W \tag{6.8a}$$

$$A_{eff} = 13.531\,mm^2 \tag{6.8b}$$

$$T = 3300\,K \tag{6.8c}$$

$$\Rightarrow L_{max} = 0.452\,Wmm^{-2}sr^{-1} \tag{6.8d}$$

The radiant power L_{max} refers to the full spectrum and is further reduced by the monochromator before it reaches the final sample grating. The targeted spectral resolution of the monochromator is chosen equal to the spectral width of the calibration laser diode ($\lambda = 636.5\,nm$, $\triangle\lambda = 2.2\,nm$). This determines the size of the exit slit S3 as derived in the following:

The mirrors (M1 and M4) of the monochromator with a focal length of $f_M = 220\,mm$ collimate the light onto the diffraction gratings. With a grating period of $\Lambda = 833.3\,nm$ (lattice constant: $1\,mm/1200$) the gratings then diffracts light into the first diffraction order ($m = 1$) under an angle of $\beta = \text{asin}\,(m\lambda/\Lambda)$ (cf. grating equation 1.37a).

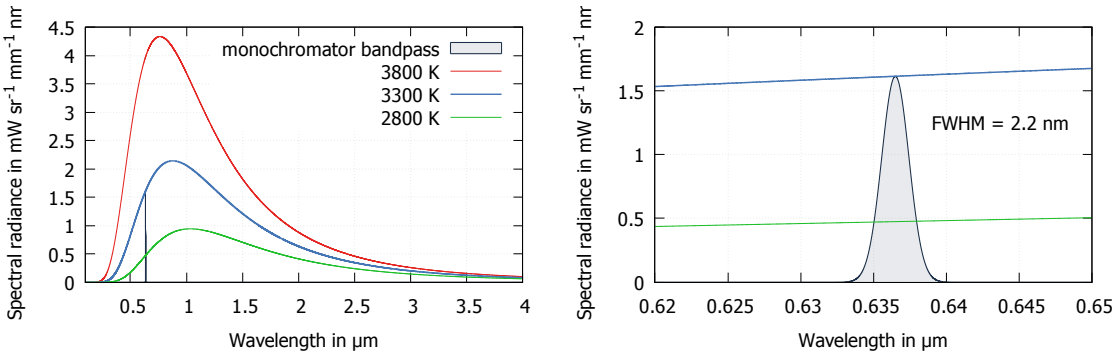
Differentiating the grating equation with respect to the diffraction angle leads to

$$\text{reciprocal angular dispersion:} \quad \gamma_\beta = \frac{d\lambda}{d\beta} = \frac{\Lambda \cos(\beta)}{m} \tag{6.9a}$$

$$\text{reciprocal linear dispersion:} \quad \gamma_x = \frac{d\lambda}{dx} \approx \frac{1}{f_M}\frac{d\beta}{d\lambda} \tag{6.9b}$$

$$\text{slit width (for } m = 1\text{) for a double monochromator:} \quad \triangle x = 2\frac{\triangle\lambda}{\gamma_x}. \tag{6.9c}$$

The factor 2 in Eq. 6.9c accounts for the doubled dispersion due to the monochromators additive configuration. Thus, in the present case, the targeted spectral width of $\triangle\lambda = 2.2\,nm$ leads to a slit width of $\triangle x = 1.8\,mm$.



(a) Black-body radiance ($\epsilon = 1$) for different temperatures according to Planck's law

(b) Spectral radiance filtered by the monochromator

Figure 6.21: Illustration of the spectral radiance of Black-bodies and output radiance of the monochromator

Assuming a Gaussian shaped output spectrum of the monochromator as illustrated in Fig. 6.21b, the output radiance is defined by the integral over the product of Planck's gray body spectrum and the Gaussian *intensity* function:

$$L = \frac{1}{2\pi}\int \underbrace{\left[\frac{2\pi hc^2}{\lambda^5}\frac{1}{e^{\frac{hc}{\lambda kT}}-1}\epsilon\right]}_{\text{Planck's law}} \underbrace{\left[e^{-\left(\frac{\lambda}{\sigma}\right)^2}\right]}_{\text{Gaussian function}} d\lambda \tag{6.10}$$

Here, h denotes Planck's constant, k is Boltzmann's constant, c is the speed of light in air and σ is derived from the spectral target width by $\sigma = \triangle\lambda / \left(2\sqrt{\ln(2)}\right)$. Again putting in the numbers leads to a total output radiance of the system of $L = 425.541\frac{\mu W}{mm^2 sr}$.

Etendue

The etendue (in $mm^2 sr$) characterizes the ability of an optical system to accept light and is defined by its illuminated surface area and the solid angle of the incident light. It is invariant as the light passes a loss-less system and is determined by the least optimized component. It can only be increased by diffusion, but never decreased without loss. Keeping this in mind, the proper imaging lenses and their positions can be derived from the fixed monochromator parameters matching the numerical aperture (NA) of the components.

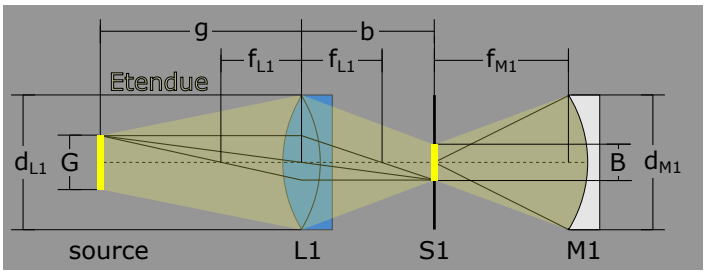


Figure 6.22: Sketch of the entrance optics

With a focal length of $f_{M1} = 220\,mm$ and a diameter of $d_{M1} = 60\,mm$ the first collimating mirror (M1, cf. Fig. 6.22) of the monochromator defines the numeric aperture of the device as $\mathrm{NA}_{M1} = \sin\left(\mathrm{atan}\left(d_{M1}/2f_{M1}\right)\right) = 0.135$. It also determines the properties of the entrance optics, which consists of a Linos microbank achromat L1 with a diameter of $d_L = 30\,mm$ that is placed $b = d_L/\left(2\tan\left(\mathrm{asin}\left(\mathrm{NA_{M1}}\right)\right)\right) = 110\,mm$ in distance to the entrance slit S1. The thin lens formula $1/g + 1/b = 1/f_L$ further determines a distance of $g = 293.33\,mm$ between the achromat and the light source at a chosen focal length of $f_L = 80\,mm$. The same optics (cf. Fig. 6.18f) is used in reverse order on the other side of the monochromator to restore the image of the source at the grating plane (cf. Fig. 6.18g). The corresponding magnification is $\mathrm{m} = b/g = B/G = 0.375$, demagnifying the coiled tungsten filament of the halogen lamp on to the entrance slit S1. The filament size is specified with $A = A_x \times A_y = 2.3\,mm \times 4.2\,mm = 9.66\,mm^2$. However, due to the long legs of the coiled filament, the effective size is increased in vertical dimension to $A_{eff} = A'_x \times A_y = 3.222\,mm \times 4.2\,mm = 13.532\,mm^2$ in accordance to Eq. 6.8b.

According to Eq. A.52 the maximal etendue of a flat surface A_{eff} radiating in the full hemisphere in air is given by

$$E_{max} = 2\pi A_{eff}. \tag{6.11}$$

The etendue is reduced by the aperture of the circular achromat L1 but should be equal on both sides of the entrance slit S1 due to the performed mode matching - and the fact that the slit size matches the size of the demagnified source image. Using the definitions in App. A.7

the reduced etendue between the circular achromat L1 and the rectangular entrance slit S1 can be expressed by

$$E_a = \int_0^{\frac{d_L}{2}} \int_0^{2\pi} \int_{-m\frac{A_y}{2}}^{m\frac{A_y}{2}} \int_{-m\frac{A'_x}{2}}^{m\frac{A'_x}{2}} \rho \left(\frac{\cos\left(\theta_{L1,S1}\right)}{l_{L1\to S1}} \right)^2 dx\, dy\, d\phi\, d\rho. \tag{6.12}$$

On the right side of the rectangular entrance slit (inside the monochromator) and before the circular mirror M1, the etendue reads as

$$E_b = \int_{-m\frac{A_y}{2}}^{m\frac{A_y}{2}} \int_{-m\frac{A'_x}{2}}^{m\frac{A'_x}{2}} \int_0^{\frac{d_{M1}}{2}} \int_0^{2\pi} \rho \left(\frac{\cos\left(\theta_{S1,M1}\right)}{l_{S1\to M1}} \right)^2 d\phi\, d\rho\, dx\, dy. \tag{6.13}$$

Putting in numbers, Eqs. 6.12 and 6.13 gives $E_a = 0.109123\,mm^2$ and $E_b = 0.109127\,mm^2$, which is effectively equal and thereby verifies the correct theoretical NA matching. Since also the monochromator preserves the etendue and the output optics mimics the entrance optic in reverse order, no significant additional loss of the geometric extend is expected after the first aperture of L1. Consequently etendue should be preserved over the entire residual system.

6.5 Measurement and conclusion

Only two more factors have to be taken into account, which are the gain in radiance power due to the rear reflector of the lamp and the losses due to absorption, reflection and diffusion at the different optical elements. The influence of the rear reflector was simply measured with the Newport powermeter at the position of the sampling grating. $14\,\mu W$ of as the radiant output power were measured with the rear reflector in place. Without it, only $11\,\mu W$ was measured, leading to a gain of $27.27\,\%$ ($\eta_B = 1.2727$).
In order to estimate the system losses, the collimated laser diode (at wavelength $\lambda = 636.5\,nm$) was used and measured in the focus of L1 before the entrance slit S1 and again behind the exit slit S3. Thereby the influences of the two achromats L1 and L2 were ignored. $47\,\mu W$ were measured before the monochromator and $11\,\mu W$ were measured behind it leading to a system efficiency of $\eta_S \approx 23.4\,\%$. Putting all calculated results together as

$$P_{out} = EL\eta_B\eta_S \tag{6.14}$$

gives an estimated output power of $13.83\,\mu W$ for the calibration wavelength of $\lambda = 636.5\,nm$ compared to $14\,\mu W$ output power, which was actually measured with the Newport powermeter.

The FWHM of the output spectrum was measured with the mut Spectrometer leading to 2.6 nm (cf. Fig. 6.23) compared to a calculated FWHM of 2.2 nm by Eq. 6.9c. The difference can be addressed to an instrumental spectral broadening.

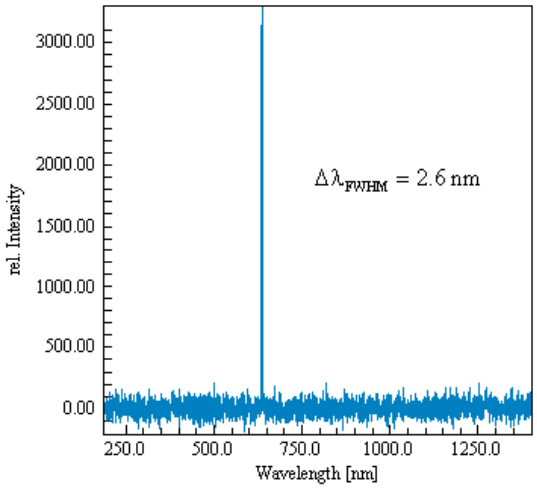


Figure 6.23: Measured spectrum of the tungsten halogen lamp filtered by the monochromator

So in conclusion the monochromator setup is able to reach an output power of 14 μW at a wavelength of $\lambda = 636.5\,nm$ and a spectral width of 2.6 nm. This is well within the order of magnitude of the tunable laser with 30 μW, which was used in the measurements in Sec. 6.3 - especially considering the smaller spectral width compared to 13.29 nm of the tunable laser. The setup also allows for variable spectral width by adjusting the exit slit width and trading spectral resolution for higher output power. A further increase of the output power could be reached by using larger achromats for the entrance and exit optics. This, however, would also require a revision of the mounting. Finally, the monochromator is even easier to handle than the delicate tunable laser, where frequent mode hopping was another reoccurring problem in the previous measurements.
Despite several minor assumptions and simplifications, the calculation shows very high accuracy, which indicates that the main influencing parameters of the system are under control. Thus, the system should be useful for future resonant measurements. The results of this project were published at an earlier stage in [3].

Summary

In this thesis a new and efficient way was developed to analyze the rigorous diffraction grating problem for incident waves beyond the mere plane wave input. The method extends the rigorous coupled-wave analysis (RCWA) but does not increase the complexity of the algorithm. As long as the incident field can be described by a finite number of plane waves, which corresponds to the number of Fourier modes used in the RCWA algorithm, the input field can be chosen arbitrarily. The new method is named localized input field RCWA (LIF-RCWA) and can be applied to various diffraction problems, which depend on finite incident beams and specially shaped wave fronts. This was demonstrated in multiple examples. The discussion included the modeling of a Gaussian beam and its influence on the diffraction grating problem. It also included the emulation of the readout process of an optical storage medium and the rigorous analysis of single- and multi-mode waveguides. Both the coupling as well as the formation and propagation of distinct waveguide modes have been demonstrated, showing excellent agreement with the analytical model. Furthermore, the modeling of a point light source was presented together with the new ability to model various forms of polarization like azimuthal and radial polarization.

From the beginning, the method was intended for the three-dimensional grating problem, which includes gratings that are periodically structured in two directions and illuminated in oblique incidence. In this general case a distinction in the two typical polarization types TE and TM is unprofitable, since Maxwell's equations do not separate as they would do in classical mounting of a one-dimensional grating. Therefore, the RCWA is slightly reformulated in Chap. 2 to work in a global Cartesian coordinate system instead of the usually used and less convenient local coordinate TE-/TM-systems of the individual diffraction orders. This not only simplifies the use of the LIF extension, but also improves clarity for the three-dimensional RCWA algorithm.

The development of the LIF-RCWA also revealed an inconsistent truncation scheme for the grating permittivity in the standard RCWA. In numerical calculations a truncation of infinite Fourier series expansions is inevitable and is usually associated with a bandwidth limitation. In the truncation process of the standard RCWA, however, some frequencies of the permittivity are underrepresented, while other frequencies beyond the given bandwidth limit are falsely accounted for. As a result, the Fourier coefficients of the permittivity do not represent a proper bandwidth limited grating. The LIF-RCWA resolves this issue by the use of a cyclic definition of the convolution matrix, which corresponds to a discrete Fourier transformation (or Fast Fourier Transform) with the same number of sampling points in space and frequency domain. It was demonstrated that this approach does not show the critical Gibbs phenomenon (on its discrete grid) and that the inverse rule, which is usually used to cure the bad convergence behavior due to this phenomenon, is in fact a neutral and redundant operation for the LIF-RCWA. This way, the convergence speed of the LIF-RCWA is faster in the case of classical TM polarization compared to the standard RCWA, but slower in the TE case - due to the decreased spatial sampling in the LIF-RCWA. Applying the inverse rule to the standard RCWA also leads to a superior convergence speed in the TM case. However, for the costs of the novel introduced consistent truncation scheme, the inverse rule can also

be applied effectively to the LIF-approach, achieving the same improved convergence speed, while maintaining the ability to treat non-plane wave input fields.
In Chap. 4 the exact calculation of the near-fields in the RCWA was investigated and verified against its compliance with Maxwell's continuity conditions as well as its similarity to a converged near-field of the standard method. Since the finite Fourier expansion struggles to properly follow the jump discontinuity of a binary grating, Lalanne and Jurek[67] suggested to derive the discontinuous field components from the continuous ones and factor in the non-bandlimited permittivity in position space. However, this involves the intermixture of bandlimited and non-bandlimited quantities, which leads to high peaks of the field amplitudes at the grating boundaries. A least square comparison between the different field-calculation methods demonstrate consistency between the convergence speeds of the near-fields and their corresponding diffraction coefficients, which also indicates a consistent calculation of the near fields. At this point, however, the influence of an inconsistent truncation scheme still remains unclear.
Chap. 5 compares different methods for the calculation of absorption in grating structures and especially investigates the method of local absorption by Brenner[13]. The latter method is also used to verify the conservation of energy of the RCWA in lossy media. While the simple relation $1-R-T \stackrel{!}{=} 0$ is usually used to confirm energy conservation in non-absorbing structures, for absorbing structures, the same relation can only be used to derive a power deficit. However, until now it remained unclear how much of this power deficit could be addressed to the actual absorbed power in the structure. Using the theory of local absorption, the electromagnetic near-fields, which were derived in the previous chapter, can be used to actually calculate the absorption. A comparison between the integrated local absorption and the power deficit finally allows a verification of the conservation of energy in lossy media. The result is that the standard RCWA does in fact not fully conserve energy for low mode counts. The LIF-RCWA, on the other hand, shows accurate agreement between absorption and the power deficit for any order of truncation. In fact, the integrated local absorption, which depends on the sampling of the field, converges towards the power deficit for an increasing z-sampling, while an optimal sampling in tangential direction is even predetermined. Both is not true for the standard RCWA and any other variant that uses the non-cyclic Toeplitz matrix to model the material parameters. The deviation between absorption and power deficit becomes even larger, when applying the inverse rule. On the other hand, the use of an appropriate apodization function, which does not cure the Gibbs problem but suppresses higher frequencies in the grating permittivity, decreases the deviation. All these facts indicate that the inconsistent truncation scheme in the non-cyclic Toeplitz matrix of the material properties is the reason for the observed discrepancies. It was even shown that the conservation of energy in the non-absorbing case is not sufficient indicator for the correctness of the calculation. The erroneous truncation is used twice when the inverse rule is applied, is reduced with apodization and corrected by the LIF approach. Using the field calculation method of Lalanne and Jurek[67] also negatively affects the energy balance. By applying a non-band-limited permittivity function to an otherwise bandwidth-limited calculation, the energy of the incident wave is redistributed over infinite frequencies, while only a finite subset is considered in the further calculation. This leads to an energy loss that is not accounted for in the analysis.
In Chap. 6 the RCWA and the theory of local absorption were used to develop a photodetector as an active optical element, which is fully compatible with the SOI-CMOS process. This

means that the detector can be built using the same materials, structure sizes and processes as any electronic device in the technology and does not require any post-processing steps for the integration into a standard chip design. The SOI-CMOS process is a standard process for electronic chips and was even shown to support some passive optical functions. It is therefore particularly suitable for this attempt. The small thickness of the photoactive silicon layer has been compensated by utilizing existing structures of the SOI-CMOS-process as a grating structure to locally enhance the intensity and especially local absorption in the photoactive layer, which finally generates a measurable photocurrent. This way, the absorption in the photoactive layer was increased from 2 % to 43 % and 86 % in simulations depending on the design - for one type of polarization. The fabricated detector was measured and the sensitivity as well as the wavelength dependency was verified in experiment. However, different deviations of the design parameters during the fabrication process led to a shift of the sensitivity maximum from $850\,nm$ to $830\,nm$, which could only partially be explained.
With a higher degree of freedom to structure the silicon trenches in the CMOS-process, also two-dimensional designs would have been possible. In simulation, those were shown to reach full polarization independence and thereby achieve an increase of efficiency by another 100 % in case of unpolarized incident light.
A final study investigated the suitability of a monochromator as a wavelength-tunable light source to illuminate and test diffractive structures. Using an off-the-shelf tungsten halogen lamp with a power rating of $100\,W$, $14\,\mu W$ of output power could be reached using a spectral width of only $2.2\,nm$ at a wavelength of $636.5\,nm$. This is within the order of magnitude of the illumination power ($30\,\mu W$) from the tunable laser source, which was used for the measurements of the CMOS-detector - especially considering that the spectral width of a monochromator was significantly smaller ($2.2\,nm$ compared to $13.29\,nm$) and can be easily adjusted in the monochromator setup for higher power output. Thus, the power is shown to be sufficient for following tests of diffractive structures.
Within the context of this work also a simulation software was developed, which was used to simulate all RCWA-related simulations presented in this thesis. Some details are discussed in the appendix B.

In conclusion, this work contributes to the fields of rigorous electromagnetic simulation and micro-optics integration. Both have significant relevance in todays research and industry and will probably continue to grow further at an enormous speed.
Even though the time complexity of the RCWA is a limiting factor, the method has become one of the most important instruments for rigorous simulation due to its wide generality. Thus, as in the previous decades, further improvements of the method and the easy access to massive parallel computing power will allow the analysis of significantly more complex problems especially in the three-dimensional regime. This will certainly open the door to new discoveries and applications in the near future. The integration of optically active components like a photodetector into the SOI CMOS process will hopefully enable a cheap production of opto-electronic hybrids that enrich conventional electronic mass market products with the advances of modern photoncis.

Appendix A

Formulae and definitions

A.1 The normalized wave vector

The normalized wave vector $\mathbf{s}$ is defined in spherical coordinated (radius = 1) as

$$\mathbf{s} = \begin{pmatrix} \sin(\theta)\sin(\phi) \\ \sin(\theta)\cos(\phi) \\ \cos(\theta) \end{pmatrix}, \tag{A.1}$$

where θ denotes the angle of incidence and ϕ denotes the azimuth. Together they describe the only propagation direction of a plane wave towards the point of origin. The angles are depicted in Fig. 2.1 for a plane incident wave of a typical diffraction grating problem in conical mounting. The length of the wave vector $\mathbf{k}$ is defined by the wavelength λ and the surrounding refractive index n (cf. Eq. 1.14).

A.2 The complex permittivity and the refractive index

The diacritic hat will indicate the complex nature of a quantity in the following. n and ϵ_r are the real parts of the refractive index and the permittivity, κ and ϵ_i are the corresponding imaginary parts. Thus, their definitions read as

$$\hat{n} = n + i\kappa \quad \text{and} \tag{A.2}$$

$$\hat{\epsilon} = \epsilon_r + i\epsilon_i \tag{A.3}$$

and they are connected by the Maxwell relation

$$\hat{n} = \sqrt{\epsilon\mu}. \tag{A.4}$$

For $\mu = 1$, the complex refractive index can be expressed as a function of the complex permittivity

$$\hat{\epsilon}\left(\hat{n}\right) = \left(n^2 + \kappa^2\right) + i2n\kappa \tag{A.5}$$

and visa versa

$$\hat{n}\left(\hat{\epsilon}\right) = \sqrt{\frac{r+\epsilon_r}{2}} + i\sqrt{\frac{r-\epsilon_r}{2}} sgn\left(\epsilon_i\right) \tag{A.6}$$

with

$$r = \sqrt{\epsilon_r^2 + \epsilon_i^2}. \tag{A.7}$$

A.3 Physical fields and time averaging

Between a complex field $\mathbf{F}$ with harmonic time dependence and a physical field $\mathbf{F_{phys}}$ the following equations holds:

$$\mathbf{F_{phys}} = Re(\mathbf{F}) = \frac{1}{2}\left(\mathbf{F}(\mathbf{r})e^{-i\omega t} + \mathbf{F}^*(\mathbf{r})e^{+i\omega t}\right). \tag{A.8}$$

Besides, time averaging is defined as

$$\langle \mathbf{G} \rangle = \lim_{T\to\infty} \frac{1}{T} \int_0^T \mathbf{G}(t)\, dt. \tag{A.9}$$

Example

Both relations can be used to derive the physical and temporal averaged Poynting vector from its complex mathematical description:

$$\mathbf{S_{phys}} = \mathbf{E_{phys}} \times \mathbf{H_{phys}} \tag{A.10}$$

$$\mathbf{S_{phys}} = \mathrm{Re}\left(\mathbf{E}\right) \times \mathrm{Re}\left(\mathbf{H}\right) \tag{A.11}$$

$$= \frac{1}{2}\left(\mathbf{E}(\mathbf{r})e^{-i\omega t} + \mathbf{E}^*(\mathbf{r})e^{+i\omega t}\right) \times \frac{1}{2}\left(\mathbf{H}(\mathbf{r})e^{-i\omega t} + \mathbf{H}^*(\mathbf{r})e^{+i\omega t}\right) \tag{A.12}$$

$$= \frac{1}{2}\mathrm{Re}\left(\mathbf{E}(\mathbf{r}) \times \mathbf{H}^*(\mathbf{r})\right) + \frac{1}{2}\mathrm{Re}\left(\mathbf{E}(\mathbf{r}) \times \mathbf{H}(\mathbf{r})e^{-2i\omega t}\right) \tag{A.13}$$

$$\langle \mathbf{S_{phys}} \rangle = \frac{1}{2} \operatorname{Re} \left(\mathbf{E}(\mathbf{r}) \times \mathbf{H}^*(\mathbf{r}) \right) + \underbrace{\lim_{T \to \infty} \frac{1}{T} \int_0^T \frac{1}{2} \operatorname{Re} \left(\mathbf{E}(\mathbf{r}) \times \mathbf{H}(\mathbf{r}) e^{-2i\omega t} \right) \, dt}_{=0} \tag{A.14}$$

$$= \frac{1}{2} \operatorname{Re} \left(\mathbf{E}(\mathbf{r}) \times \mathbf{H}^*(\mathbf{r}) \right) \tag{A.15}$$

$$= \frac{1}{2} \operatorname{Re} \left(\mathbf{E}(\mathbf{r}) e^{-i\omega t} \times \mathbf{H}^*(\mathbf{r}) e^{+i\omega t} \right) \tag{A.16}$$

$$= \frac{1}{2} \operatorname{Re} \left(\mathbf{E} \times \mathbf{H}^* \right) \tag{A.17}$$

A.4 The C matrix

The rigorous coupled-wave analysis solves the diffraction problem by matching the tangential field components of the electric and magnetic fields. In this work the **C** matrix is used to convert the tangential components of the Fourier harmonics of an electric field to those of the corresponding magnetic field. The **C** matrix can easily be derived from the second Maxwell equation for harmonic fields (cf. Eq. 1.8b):

$$\nabla \times \mathbf{E} = i\omega \mathbf{B} \tag{A.18}$$

$$\Rightarrow \nabla \times \mathbf{E} = i\omega\mu\mu_0 \mathbf{H} \tag{A.19}$$

$$\Rightarrow \mathbf{k} \times \mathbf{E} = k_0 \mu Z_0 \mathbf{H} \tag{A.20}$$

For the tangential field components of the magnetic field, this leads to

$$\begin{pmatrix} k_y E_z - k_z E_y \\ k_z E_x - k_x E_z \end{pmatrix} = k_0 \mu Z_0 \begin{pmatrix} H_x \\ H_y \end{pmatrix}. \tag{A.21}$$

The normal (z-)component of the electric field can be substituted according to transversal property of the field (cf. Eq. 1.25):

$$E_z = -\frac{k_x E_x + k_y E_y}{k_z} \quad \text{for } k_z \neq 0 \tag{A.22}$$

$$\Rightarrow \begin{pmatrix} H_x \\ H_y \end{pmatrix} = \frac{1}{\mu k_z k_0 Z_0} \begin{pmatrix} -k_x k_y & -\left(k_y^2 + k_z^2\right) \\ k_x^2 + k_z^2 & k_x k_y \end{pmatrix} \begin{pmatrix} E_x \\ E_y \end{pmatrix}. \tag{A.23}$$

Finally, in order to convert the vectors of Fourier harmonics ($\mathbf{U}_\perp = \mathbf{C}\cdot\mathbf{S}_\perp$), field components and wave vectors become vector quantities and the C matrix can be defined as:

$$\mathbf{C}_l := \frac{1}{\mu k_0 Z_0} \begin{bmatrix} -\operatorname{diag}_q\left(\frac{k_{x;q}k_{y;q}}{k_{z;l;q}}\right) & -\operatorname{diag}_q\left(\frac{k_{y;q}^2 + k_{z;l;q}^2}{k_{z;l;q}}\right) \\ \operatorname{diag}_q\left(\frac{k_{x;q}^2 + k_{z;l;q}^2}{k_{z;l;q}}\right) & \operatorname{diag}_q\left(\frac{k_{x;q}k_{y;q}}{k_{z;l;q}}\right) \end{bmatrix} \quad \text{for } k_{z;l;q} \neq 0. \tag{A.24}$$

The operator $\operatorname{diag}_q(v_q)$ converts all elements v_q with $q = 0\,..\,\operatorname{count}[1](\mathbf{v}) - 1$ of a vector $\mathbf{v}$ into a diagonal matrix with $\operatorname{count}(\mathbf{v}) \cdot \operatorname{count}(\mathbf{v})$ elements. The index l only refers to one of the homogeneous outer regions $l = I$ or $l = III$. For the special case of $k_{z;j;q} = 0$ (cf. cut-off frequency, Eq. 1.18b) the matrix $\mathbf{C}$ is not defined, since H_x and H_y depends on E_z exclusively, which cannot be derived from the tangential components of the electric field. From an implementation standpoint, the identity $k_x^2 + k_z^2 = k^2 - k_y^2$ is not useful in an inhomogeneous medium, since there is no single refractive index n, from which $k = nk_0$ can quickly be calculated. The individual wave vector components on the other side need to be available anyway.

A.5 The Ω matrix

In the RCWA algorithm, the $\mathbf{\Omega}$ matrix describes the relation between the electric or the magnetic field and its second derivative in a second-order differential equation $\partial^2\mathbf{F}_\perp/\partial z^2 = \mathbf{\Omega}\mathbf{F}_\perp$. As shown in Sec. 2.4.2, only one of the fields needs to be solved with an eigenvalue ansatz, while the other can then be derived. Hence, depending on the choice of the field, the $\mathbf{\Omega}$ matrix can have two different definitions, which both shall be deduced in this section for the three- and the two-dimensional case.

In the first order differential equation 2.14, the electric and the magnetic field still appear coupled:

$$\frac{\partial}{\partial z}\begin{pmatrix}\mathbf{S}_\perp \\ \mathbf{U}_\perp\end{pmatrix} = \begin{pmatrix}0 & \mathbf{M}_1 \\ \mathbf{M}_2 & 0\end{pmatrix}\begin{pmatrix}\mathbf{S}_\perp \\ \mathbf{U}_\perp\end{pmatrix}, \tag{A.25}$$

while another derivative reveals the definition of the two variants of the $\mathbf{\Omega}$ matrix (cf. Eq. 2.16) in uncoupled equations:

$$\Rightarrow \quad \partial^2\mathbf{S}_\perp/\partial z^2 - \mathbf{\Omega}_E\mathbf{S}_\perp = 0 \qquad \text{with } \mathbf{\Omega}_E := \mathbf{M}_1\mathbf{M}_2 \tag{A.26}$$

$$\Rightarrow \quad \partial^2\mathbf{U}_\perp/\partial z^2 - \mathbf{\Omega}_H\mathbf{U}_\perp = 0 \qquad \text{with } \mathbf{\Omega}_H := \mathbf{M}_2\mathbf{M}_1 \tag{A.27}$$

[1] The operator $\operatorname{count}(\mathbf{v})$ gives the number of elements of the vetor $\mathbf{v}$.

A.5.1 The Ω matrix for the electric field

$$\boldsymbol{\Omega}_E^{3D} = \mathbf{M}_1\mathbf{M}_2 \tag{A.28}$$

$$= -k_0^2 \begin{bmatrix} \mathbf{K}_x[\![\epsilon]\!]^{-1}\mathbf{K}_y & -\mathbf{K}_x[\![\epsilon]\!]^{-1}\mathbf{K}_x + \lceil\lfloor\mu\rfloor\rceil \\ \mathbf{K}_y[\![\epsilon]\!]^{-1}\mathbf{K}_y - \lfloor\lceil\mu\rceil\rfloor & -\mathbf{K}_y[\![\epsilon]\!]^{-1}\mathbf{K}_x \end{bmatrix} \tag{A.29}$$

$$\cdot \begin{bmatrix} -\mathbf{K}_x[\![\mu]\!]^{-1}\mathbf{K}_y & \mathbf{K}_x[\![\mu]\!]^{-1}\mathbf{K}_x - \lceil\lfloor\epsilon\rfloor\rceil \\ -\mathbf{K}_y[\![\mu]\!]^{-1}\mathbf{K}_y + \lfloor\lceil\epsilon\rceil\rfloor & \mathbf{K}_y[\![\mu]\!]^{-1}\mathbf{K}_x \end{bmatrix} \tag{A.30}$$

$$= k_0^2 \quad \begin{pmatrix} \alpha_E & \beta_E \\ \gamma_E & \delta_E \end{pmatrix} \tag{A.31}$$

with

$$\alpha_E = \mathbf{K}_x[\![\epsilon]\!]^{-1}\mathbf{K}_x\lfloor\lceil\epsilon\rceil\rfloor + \lceil\lfloor\mu\rfloor\rceil\mathbf{K}_y[\![\mu]\!]^{-1}\mathbf{K}_y - \lceil\lfloor\mu\rfloor\rceil\lfloor\lceil\epsilon\rceil\rfloor \tag{A.32}$$

$$\beta_E = \mathbf{K}_x[\![\epsilon]\!]^{-1}\mathbf{K}_y\lceil\lfloor\epsilon\rfloor\rceil - \lceil\lfloor\mu\rfloor\rceil\mathbf{K}_y[\![\mu]\!]^{-1}\mathbf{K}_x \tag{A.33}$$

$$\gamma_E = \mathbf{K}_y[\![\epsilon]\!]^{-1}\mathbf{K}_x\lfloor\lceil\epsilon\rceil\rfloor - \lfloor\lceil\mu\rceil\rfloor\mathbf{K}_x[\![\mu]\!]^{-1}\mathbf{K}_y \tag{A.34}$$

$$\delta_E = \mathbf{K}_y[\![\epsilon]\!]^{-1}\mathbf{K}_y\lceil\lfloor\epsilon\rfloor\rceil + \lfloor\lceil\mu\rceil\rfloor\mathbf{K}_x[\![\mu]\!]^{-1}\mathbf{K}_x - \lfloor\lceil\mu\rceil\rfloor\lceil\lfloor\epsilon\rfloor\rceil \tag{A.35}$$

Considering only one-dimensional gratings in non-conical mounting (referred to as the 2D case) $\mathbf{K}_y$ disappears and the diffraction problem is usually separated into two cases of polarization, namely TE and TM.

For TE polarization S_x and U_y are equal to zero and the $\boldsymbol{\Omega}$ matrix simplifies to

$$\boldsymbol{\Omega}_{E,TE}^{2D} = k_0^2[1/\mu]^{-1}\left(\mathbf{K}_x[\mu]^{-1}\mathbf{K}_x - [\epsilon]\right) \tag{A.36}$$

and for TM polarization with $S_y = U_x = 0$ to

$$\boldsymbol{\Omega}_{E,TM}^{2D} = k_0^2\left(\mathbf{K}_x[\epsilon]^{-1}\mathbf{K}_x - [\mu]\right)[1/\epsilon]^{-1} \tag{A.37}$$

A.5.2 The Ω matrix for the magnetic field

In order to calculate the magnetic counterpart of the $\mathbf{\Omega}$ matrix used in Sec. 2.16, only the order of the matrices $\mathbf{M}_1$ and $\mathbf{M}_2$ has to be reversed. This example may serve as a reference, as it is used in many implementations [88, 91] especially for one-dimensional gratings.

$$\mathbf{\Omega}_H^{3D} = \mathbf{M}_2\mathbf{M}_1 \tag{A.38}$$

$$= -k_0^2 \begin{bmatrix} -\mathbf{K}_x[\![\mu]\!]^{-1}\mathbf{K}_y & \mathbf{K}_x[\![\mu]\!]^{-1}\mathbf{K}_x - \lceil\lfloor\epsilon\rfloor\rceil \\ -\mathbf{K}_y[\![\mu]\!]^{-1}\mathbf{K}_y + \lfloor\lceil\epsilon\rceil\rfloor & \mathbf{K}_y[\![\mu]\!]^{-1}\mathbf{K}_x \end{bmatrix} \tag{A.39}$$

$$\cdot \begin{bmatrix} \mathbf{K}_x[\![\epsilon]\!]^{-1}\mathbf{K}_y & -\mathbf{K}_x[\![\epsilon]\!]^{-1}\mathbf{K}_x + \lceil\lfloor\mu\rfloor\rceil \\ \mathbf{K}_y[\![\epsilon]\!]^{-1}\mathbf{K}_y - \lfloor\lceil\mu\rceil\rfloor & -\mathbf{K}_y[\![\epsilon]\!]^{-1}\mathbf{K}_x \end{bmatrix}$$

$$= k_0^2 \begin{pmatrix} \alpha_H & \beta_H \\ \gamma_H & \delta_H \end{pmatrix} \tag{A.40}$$

with

$$\alpha_H = \mathbf{K}_x[\![\mu]\!]^{-1}\mathbf{K}_x\lfloor\lceil\mu\rceil\rfloor + \lceil\lfloor\epsilon\rfloor\rceil\mathbf{K}_y[\![\epsilon]\!]^{-1}\mathbf{K}_y - \lceil\lfloor\epsilon\rfloor\rceil\lfloor\lceil\mu\rceil\rfloor \tag{A.41}$$

$$\beta_H = \mathbf{K}_x[\![\mu]\!]^{-1}\mathbf{K}_y\lceil\lfloor\mu\rfloor\rceil - \lceil\lfloor\epsilon\rfloor\rceil\mathbf{K}_y[\![\epsilon]\!]^{-1}\mathbf{K}_x \tag{A.42}$$

$$\gamma_H = \mathbf{K}_y[\![\mu]\!]^{-1}\mathbf{K}_x\lfloor\lceil\mu\rceil\rfloor - \lfloor\lceil\epsilon\rceil\rfloor\mathbf{K}_x[\![\epsilon]\!]^{-1}\mathbf{K}_y \tag{A.43}$$

$$\delta_H = \mathbf{K}_y[\![\mu]\!]^{-1}\mathbf{K}_y\lceil\lfloor\mu\rfloor\rceil + \lfloor\lceil\epsilon\rceil\rfloor\mathbf{K}_x[\![\epsilon]\!]^{-1}\mathbf{K}_x - \lfloor\lceil\epsilon\rceil\rfloor\lceil\lfloor\mu\rfloor\rceil \tag{A.44}$$

And again, in the two-dimensional, non-conical case, the $\mathbf{\Omega}$ matrix simplifies to

$$\mathbf{\Omega}_{H,TE}^{2D} = k_0^2 \left(\mathbf{K}_x[\mu]^{-1}\mathbf{K}_x - [\epsilon]\right)[1/\mu]^{-1} \qquad \text{and} \tag{A.45}$$

$$\mathbf{\Omega}_{H,TM}^{2D} = k_0^2[1/\epsilon]^{-1}\left(\mathbf{K}_x[\epsilon]^{-1}\mathbf{K}_x - [\mu]\right). \tag{A.46}$$

A.6 Conversion of reflection and transmission coefficients into local TE/TM coordinate systems

In Ch. 2 the formulation of the modified standard RCWA yields reflection and transmission coefficients in a global Cartesian coordinate system. The classical standard RCWA according to Moharam and Gaylord[91], however, aims for a description in local TE/TM coordinate systems of the individual diffraction orders. The definition of the corresponding base vectors $\mathbf{e}_N$ and $\mathbf{e}_T$ for the local coordinate system was given in Eqs. 1.27a and 1.27a. However, this time one must distinguish between reflected and transmitted diffraction orders, which point in opposite z direction.

The single diffraction coefficients can then be expressed by the local base vectors as

$$\mathbf{R}_{m,n} = R_{N;m,n}\mathbf{e}_{N;m,n} + R_{T;m,n}\mathbf{e}^{I}_{T;m,n} \quad \text{and} \tag{A.47a}$$
$$\mathbf{T}_{m,n} = T_{N;m,n}\mathbf{e}_{N;m,n} + T_{T;m,n}\mathbf{e}^{III}_{T;m,n} \tag{A.47b}$$

with

$$\mathbf{e}_N = \frac{\mathbf{e}_z \times \mathbf{s}}{|\mathbf{e}_z \times \mathbf{s}|} \tag{A.48a}$$
$$\mathbf{e}^{I}_{T;m,n} = \frac{\mathbf{e}_{N;m,n} \times \mathbf{k}^{I}_{m,n}}{|\mathbf{e}_{N;m,n} \times \mathbf{k}^{I}_{m,n}|} \quad \text{and} \tag{A.48b}$$
$$\mathbf{e}^{III}_{T;m,n} = \frac{\mathbf{e}_{N;m,n} \times \mathbf{k}^{III}_{m,n}}{|\mathbf{e}_{N;m,n} \times \mathbf{k}^{III}_{m,n}|} \tag{A.48c}$$

where $\mathbf{k}^{I}_{m,n}$ and $\mathbf{k}^{III}_{m,n}$ are the wave vectors of the reflected and transmitted diffraction orders with opposite orientation in z (cf. Eq. 2.8). Therewith, the coefficients can be described in local coordinate systems:

$$R_{N;m,n} = \frac{1}{k_{\perp;m,n}} \left(R_{y;m,n}k_{x;m,n} - R_{x;m,n}k_{y;m,n}\right) \tag{A.49a}$$
$$R_{T;m,n} = \frac{k^{I}}{\gamma^{I}_{m,n}k_{\perp;m,n}} \left(R_{y;m,n}k_{y;m,n} + R_{x;m,n}k_{x;m,n}\right) \tag{A.49b}$$
$$T_{N;m,n} = \frac{1}{k_{\perp;m,n}} \left(T_{y;m,n}k_{x;m,n} - T_{x;m,n}k_{y;m,n}\right) \tag{A.49c}$$
$$T_{T;m,n} = \frac{k^{III}}{\gamma^{III}_{m,n}k_{\perp;m,n}} \left(T_{y;m,n}k_{y;m,n} + T_{x;m,n}k_{x;m,n}\right) \tag{A.49d}$$

For the special case, where the (m,n)th diffraction order is exactly normal to the grating and $k_{\perp;m,n} = 0$, the base vectors have been defined as $\mathbf{e}_N = \mathbf{e}_y$ and $\mathbf{e}_T = \mathbf{e}_x$ (cf. Sec. 1.2:polarization). Substituting those into Eqs. A.47a and A.47b simply leads to

$$R_{N;m,n} = R_{y;m,n} \tag{A.50a}$$
$$R_{T;m,n} = R_{x;m,n} \tag{A.50b}$$
$$T_{N;m,n} = T_{y;m,n} \tag{A.50c}$$
$$T_{T;m,n} = T_{x;m,n} \tag{A.50d}$$

A.7 Etendue

The etendue of a light emitting source is defined as the product of its area and the solid angle into which it radiates. Between a source S and a target T it can be described by

$$dE_S = n^2 dA_S \cos(\theta_S)\, d\Omega_T, \tag{A.51}$$

where n is the refractive index of the surrounding medium, dA_S is a surface element of the emitting source and $d\Omega_T$ is the solid angle, which holds the radiating beams. θ_S is the angle between the normal vector of the surface element and the direction to the target T from where the etendue is measured. For a solid angle, which spans a full hemisphere ($\Omega_T = 2\pi$) at $\theta_S = 0$, the maximum etendue is simply given by

$$E_{max} = 2\pi n^2 A. \tag{A.52}$$

The solid angle can be defined as the area of the segment of a unit sphere divided by the square of its radius r. For an infinitesimal area dA_T, which is tilted by θ_T relative to the origin of the sphere, a solid angle element $d\Omega_T$ can be expressed as

$$d\Omega_T = dA_T \cos(\theta_T)/r^2. \tag{A.53}$$

This leads to a differential etendue of

$$dE_S = n^2 \frac{\cos(\theta_S)\cos(\theta_T)}{l^2} dA_S dA_T \tag{A.54}$$

where l is the distance between the two surface elements and θ_S and θ_T are the angles between the normal vectors of these surface elements to their connecting line. If source and target surface are in parallel mounting, the angles θ_S and θ_T are equal and the two cosine expression reduce to $\cos^2(\theta)$.

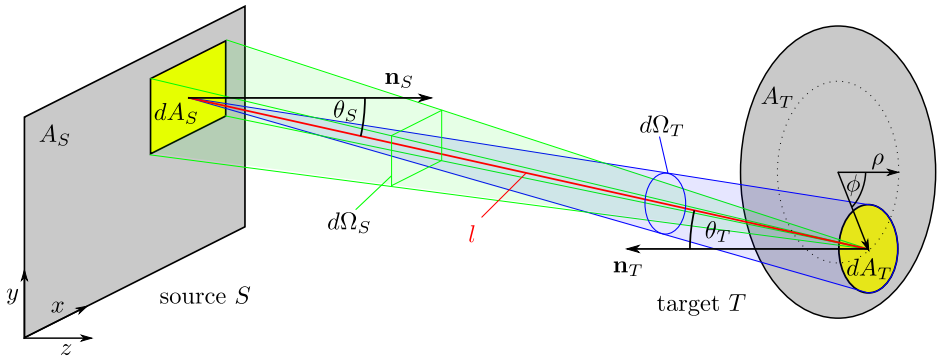


Figure A.1: Illustration of the etendue between a light emitting rectangular source and a circular target screen

If, furthermore, the source surface is defined in Cartesian coordinates and the target surface is defined in cylindric coordinates as depicted in Fig. A.1 (or the way around), then the distance $l_{S\to T}$ between dA_S at $(x, y, 0)^T$ and
dA_T at $(\rho\cos(\phi), \rho\sin(\phi), z)^T$ is

$$l_{S\to T} = \sqrt{z^2 + v^2} \tag{A.55}$$

with

$$v = \sqrt{x^2 + y^2 + \rho^2 - 2\rho\left(x\cos(\phi) + y\sin(\phi)\right)}. \tag{A.56}$$

And the etendue is

$$E = \int_{-\frac{A_y}{2}}^{\frac{A_y}{2}} \int_{-\frac{A_x}{2}}^{\frac{A_x}{2}} \int_{0}^{\frac{d}{2}} \int_{0}^{2\pi} \rho \left(\frac{n\cos(\theta_{S,T})}{l_{S\to T}}\right)^2 d\phi\, d\rho\, dx\, dy \tag{A.57}$$

with

$$A_s = A_x \cdot A_y, \tag{A.58}$$
$$A_T = \pi\,(d/2)^2 \tag{A.59}$$

and

$$\theta_{S,T} = \operatorname{atan}\left(\frac{z}{v}\right). \tag{A.60}$$

Appendix B

Simulation Software

All RCWA and LIF-RCWA simulations were performed with a custom software that was developed within the scope of this work. The source code was written in Object Pascal, which is the standard programming language at the chair of optoelectronic under Prof. Brenner. This enabled the reuse of optimized math and data visualization libraries, which are in continuous development at the chair. Object Pascal is a high level, object-oriented programming language. And despite its low popularity it is a very efficient language for the development of Windows desktop applications. Its special features include a well structured syntax and a fast one-pass compiler, which produces efficient native machine code.

The following section provides a short overview over the software architecture and a selection of the most important features and design considerations. However, it does not aim for a detailed documentation.

The goals of the simulation software were

1. ... a graphical user interface.
2. ... high flexibility in the description of the diffraction problem.
3. ... high flexibility and efficiency in the execution of parameter scans.
4. ... high flexibility in the choice of output parameters.
5. ... intelligent use of resources[1].

The goals were achieved in the same order by

1. ... the use of the Embarcadero Delphi Development Environment for rapid application development.
2. ... the implementation of interchangeable models for different light sources and gratings including one that allows a description in simple mathematical syntax.
3. ... the use of the Runtime-Type-Information (RTTI) concept, which allows access to class type information of generic objects during runtime. This way, properties of an otherwise unknown object can be read and modified.
4. ... the implementation of various output quantities (cf. list 5)
5. ... the reuse and archiving of intermediate results and the parallel execution of parameter scans.

[1] Note: The standard RCWA algorithm has a runtime complexity of $O\left(n^6\right)$ (cf. Sec. 3.18b).

B.0.1 Structure

The software's architecture follows the model-view-controller pattern in order to separate data structures from program logic and the user interface. The main components and their interplay are depicted in the data flow diagram in Fig. B.1. A user has only access to the view components, which are the `PupilEditor` (1), the `GratingEditor` (2), the `Simulator` (3) and the `Viewer` (4). The `PupilEditor` (1) and the `GratingEditor` (2) allow the user to define the incident light as well as the diffraction grating with several structured layers. Both editors are implemented using the concept of Runtime-Type-Information (RTTI). The result is a graphical user interface component (also referred to as ObjectInspector (cf. Figs. B.2, B.3 and B.4: list structure on the left side)) that can list and modify properties of a generic object. This allows a very flexible design, since the underlying models can be exchanged during runtime without any changes in the main program. And this is even true if the various models depend on different design parameters. As an example, a Gaussian beam model would need an extra parameter σ to describe the beam waist. The only requirement is that different models of the same kind produce the same standardized intermediate results for further processing. These include the Toeplitz matrix of the permittivity and the incident field vector as used in the RCWA. A complete overview is given in list 1, 2 and 3. All parameters together fully describe the diffraction problem and are summarized in a `Setup` object (cf. Fig. B.1).

PupilEditor

The `PupilEditor` provides a graphical user interface for the definition of the incident field and therefore enables the parameterization of one of the following pupil models, which were implemented:

- a `PlaneWaveModel` for standard RCWA calculations
- a `GaussianBeamModel` for finite beams
- an `AllModesModel` for perfect foci

GratingEditor

The `GratingEditor` allows a similar access to the grating representation and to the layer stack, which forms the grating. Each layer can again be defined by a different layer model. The implemented layer models include one analytical model and two FFT models. The `LayerModels` describe the x- and y-dependent permittivity function as well as the permeability function in one fundamental period of the grating layer and calculate the corresponding Fourier coefficients. The following models were implemented:

- the `AnalyticalModel` uses parametrized Fourier-pairs
- the `FFTFormulaModel` allows a basic mathematical description of a very general layer geometry.
- The `FFTBitmapModel` is able to translates a monochrome bitmap into a binary grating layer.

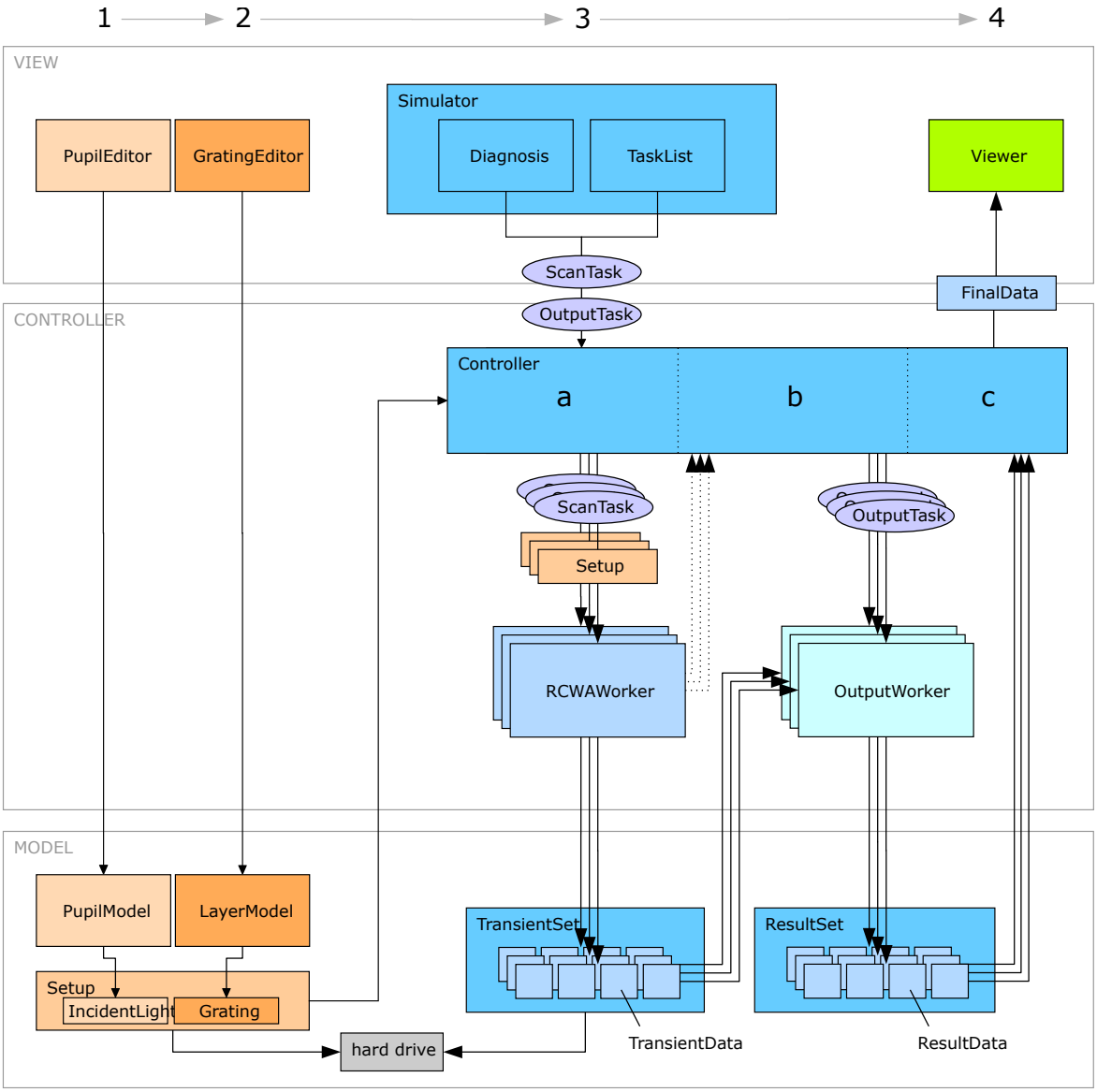


Figure B.1: Workflow diagram of the simulation program. Square boxes indicate essential modules and data structures of the architecture. Round boxes stand for messages with calculation instructions. Solid arrows show the direction of data processing and the dotted arrow lines indicate notifications.

Simulator

The main component of the `Simulator` is an editable task list (cf. Fig. B.3: list in the middle). A task defines the desired output quantity. Lists 5 and 6 show all implemented outputs ranging from the electric and magnetic field via single diffraction coefficients to local absorption. However, a task can also specify one or two `Setup` parameters, which are later varied over a certain range in order to plot their influence on a specific output quantity. An important feature of the software is a separation of these two types of tasks in `OutputTasks`

and ScanTasks. This allows the reuse of the time consuming intermediate results of a ScanTask by multiple OutputTasks. These intermediate results are called TransientData and can also be archived on a local hard drive together with the Setup data for a fast analysis with any OutputTask at a later time. Since not only a ScanTask can produce a two-dimensional output, but also an OutputTask could request a two-dimensional slice of say the electric field, certain combinations of ScanTasks and OutputTasks would lead to four-dimensional outputs. However, by convention the output is always limited to two dimensions: A scanning task can for example vary the wavelength and the grating's period over a certain range, so that an associated output task could visualize the corresponding transmission efficiencies of the setup. Alternatively, a zero-dimensional ScanTask could calculate a single diffraction problem, and an associated output task could then visualize a two-dimensional slice of the electromagnetic near-field in a given region. If only a single numeric value is requested, the simulator also offers a diagnosis tool, which saves the effort for the user to define two tasks.

Controller

The Controller has a permanent accesses to the Setup object and is notified about any changes by an implemented observer pattern. If a calculation is triggered in the Simulator interface, the Controller receives a ScanTask and an associated OutputTask. In order to speed up the calculation and to make use of the multiple cores of a modern computer, CalculationTasks and OutputTasks with a dimension higher than zero are processed by multiple threads in parallel. For this purpose, the Controller dublicates the CalculationTask and the Setup data for several RCWAWorker threads (cf. Fig. B.1). It may also change the order of a scan parameters in the two-dimensional calculation in order to move time-consuming calculations in the outer of two nested loops. Finally, additional information is added to the tasks so that every thread can work on its own local Setup at a different parameter configuration. Local copies are required, since the Setup data primarly contain the internal representation of the incident field and the grating (e.g. Toepitz matrices), which needs to be updated as soon as system parameters change. The RCWAWorkers provide the diffraction coefficients and the results of the eigenvalue problem. Together, the results are stored for each diffraction problem as so called TransientData in a global TransientSet. After completion, every RCWAWorker thread notifies the Controller, which then activates an Outputworker to further process its chunk of TransientData. Every OutputWorker receives a copy of the OutputTask and calculates the requested output quantity accordingly. The results (ResultData) are again stored in a global ResultSet. The Controller waits for the last OutputWorker to finish its calculation and then merges and rearranges partial results of all Outputworkers into a form that can be displayed by the Viewer (cf. FinalData in Fig. B.1). As mentioned above, instead of forwarding the TransientData to the Outputorkers, the Controller can also store the results on a local hard drive for a later analysis.

Viewer

The purpose of the `Viewer` is simply to visualize the `FinalData`, which is basically a zero-, one- or two-dimensional array of double-precision floating-point numbers, depending on the requested output. The data can be visualized as an array of numbers or as a one- or two-dimensional graphical plot.

Mathparser

The `MathParser` was developed to allow a mathematical description of the permittivity function $\epsilon(x, y)$ and permeability function $\mu(x, y)$ in simple syntax. The parser can be accessed via the `FFTFormulaModel`, which provides a text input field for the user (cf. Fig. B.2: bottom region above the log window). The parser receives a description in infix notation, internally translates it into reverse polish notation (RPN) and then evaluates the mathematical expression on a discrete grid for a fundamental period of a grating layer. A typical valid description reads as

infix notation (user input)

$$1.1 \,\&\, \mathrm{rect}\,(0, 0, f \cdot Px, Py) \cdot (1.5 \,\hat{}\, 2) \tag{B.1}$$

reverse polish notation

$$\&\ 1.1 \cdot \mathrm{rect}\ 0\ 0 \cdot Px\ f\ Py\ \hat{}\ 1.5\ 2 \tag{B.2}$$

The example describes a one-dimensional rect function in the center of the fundamental grating period Px with a duty cycle of f and an amplitude of $\epsilon_a = 2.25$ in a surrounding medium with $\epsilon_b = 1.1$. Thus, the parsers can process rudimentary math operations like $+$, $-$, $\cdot$, $/$ and $\hat{}$. It understands nested parentheses, has access to some setup properties like the grating periods Px and Py, allows the definition of user defined variables (such as f) (see also Fig. B.2: bottom left of the ObjectInspector) and finally can process predefined functions like a rect function with multiple arguments (here: left-shift, top-shift, width and height). The special & operator always returns its right hand argument if it is different to zero and its left hand argument otherwise. The rect function is defined in a way that it returns the value 1 if the processed grip point falls within its defined borders and the value 0 otherwise. So, if the parser evaluates the current example of Eq. B.1, it will return the value 1.1 if the grid point falls outside the borders of the rect and $1 \cdot 1.5 \,\hat{}\, 2 = 2.25$, if the processed grid point falls withing the borders. (The translation from the infix notation in Eq. B.1 to the RPN in Eq. B.2 is only possible, because the number of parameters of the rect-function is made available to the parser during its definition.) Other implemented functions are for example sin, cos, sqrt, exp.

B.0.2 Setup properties

The following listing shows all parameters, which define a diffraction grating problem in the RCWA (cf. Chap. 2) and furthermore all `Setup` properties that are accessible either directly or indirectly by the user in order to define the problem using the described software.
All parameters with the exception of the `<internal data>` are directly accessible by the user via the `PupilEditor` or the `GratingEditor`. The `<internal data>` are generated by the individual model. All blue properties are required by the RCWA. The red color indicates additinal parameters, which are provided by a specific model to support the description of the grating problem.

incident field properties **[1]**

- λ (wavelength)
- θ, ϕ, ψ (incident, azimuth and polarization angle)
- M_0, N_0 (truncation order)
- <additional model data> (e.g. σ, x-shift, y-shift, z-shift)
- <internal data>
  - $\mathbf{L}$ (incident electric field vector)

grating properties **[2]**

- n_I, n_{III} (surrounding refractive indices)
- P_x, P_y (grating periods)
- <layer list>

layer properties **[3]**

- d (layer thickness)
- <user defined variables> (they may help to define $\epsilon(x,y)$ and $\mu(x,y)$)
- <additional model data> (e.g. sampling N_x, N_y for FFT models)
- <internal data>
  - $[\![\epsilon]\!]$, $\lceil\lfloor\epsilon\rfloor\rceil$, $\lfloor\lceil\epsilon\rceil\rfloor$ (Toeplitz matrix of the permittivity ...
  - $[\![\mu]\!]$, $\lceil\lfloor\mu\rfloor\rceil$, $\lfloor\lceil\mu\rceil\rfloor$ (..., permeability and inverse Li matrices)

All configuration parameters of the software including the description of the diffraction problem (blue properties, see below) and the task list is stored in a human readable ini-file, which can even be passed to the software as a command line argument for batch processing.

B.0.3 Transient data

transient data **[1]**

- $\mathbf{L}$, $\mathbf{R}$, $\mathbf{T}$
- $\mathbf{r}_l$, $\mathbf{t}_l$ per layer l
- $\mathbf{V}_l$, $\mathbf{W}_l$, $\mathbf{Q}_l$ per layer l

B.0.4 Output parameter

output parameters [4]

- `permittivity / permeability`
  - `design / reconstruction`
    - `real part / imaginary part`
- `E-field / H-field`
  - x `/` y `/` z
    - `real part / imaginary part / absolute value / phase`
  - $|E|$, $|E|^2$
- `coefficients`
  - `reflection / transmission`
    - x `/` y `/` z
      - `all modes` m `/ specific mode` $m = m_0$
      - `all modes` n `/ specific mode` $n = n_0$
    - $|E|$, $|E|^2$
- `diffraction order efficiencies`
  - `reflection / transmission`
    - `all modes` m `/ specific mode` $m = m_0$
    - `all modes` n `/ specific mode` $n = n_0$
- `global efficiency (summed value)`
  - `reflection / transmission`
- `local absorption`
  - `design / reconstruction`
    - $x = ? .. ?, N_x = ?$
    - $y = ? .. ?, N_y = ?$
    - $z = ? .. ?, N_z = ?$

output field scan parameters [5]

- $x = ? .. ?, N_x = ?$
- $y = ? .. ?, N_y = ?$
- $z = ? .. ?, N_z = ?$

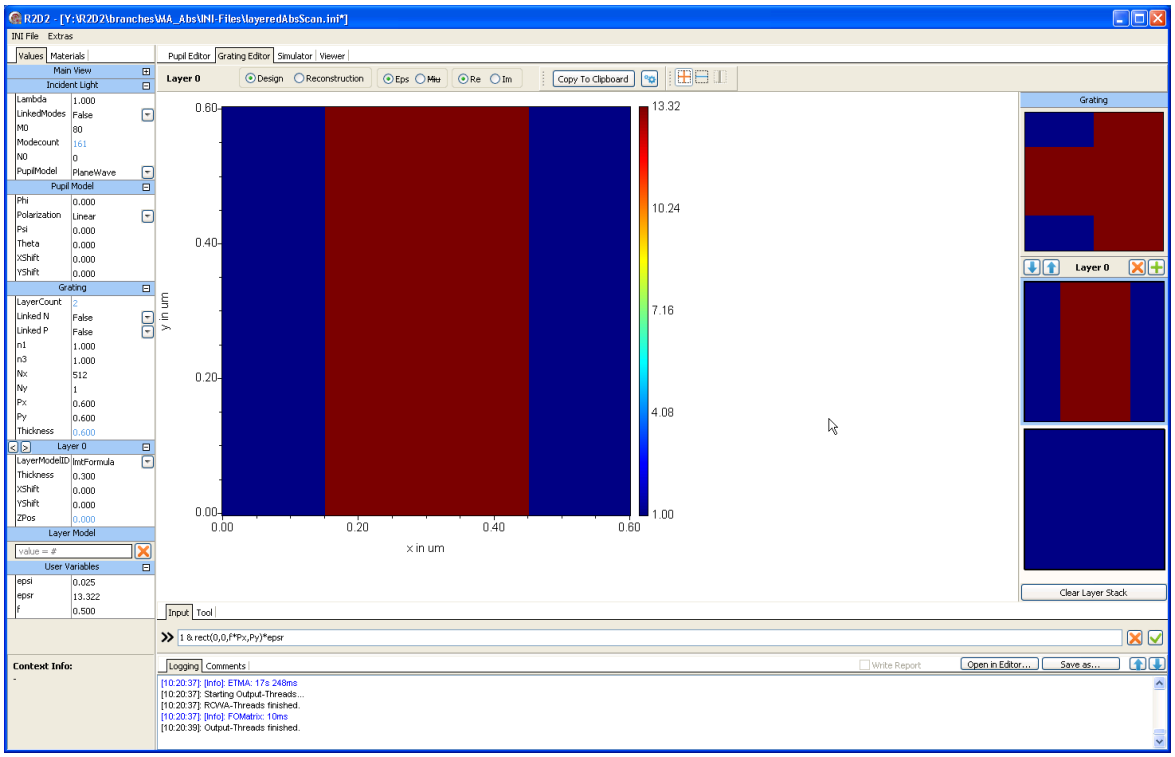


Figure B.2: Screenshot of the `GratingEditor`. A text input field at the bottom of the screen allows a simple mathematical description of the grating structure

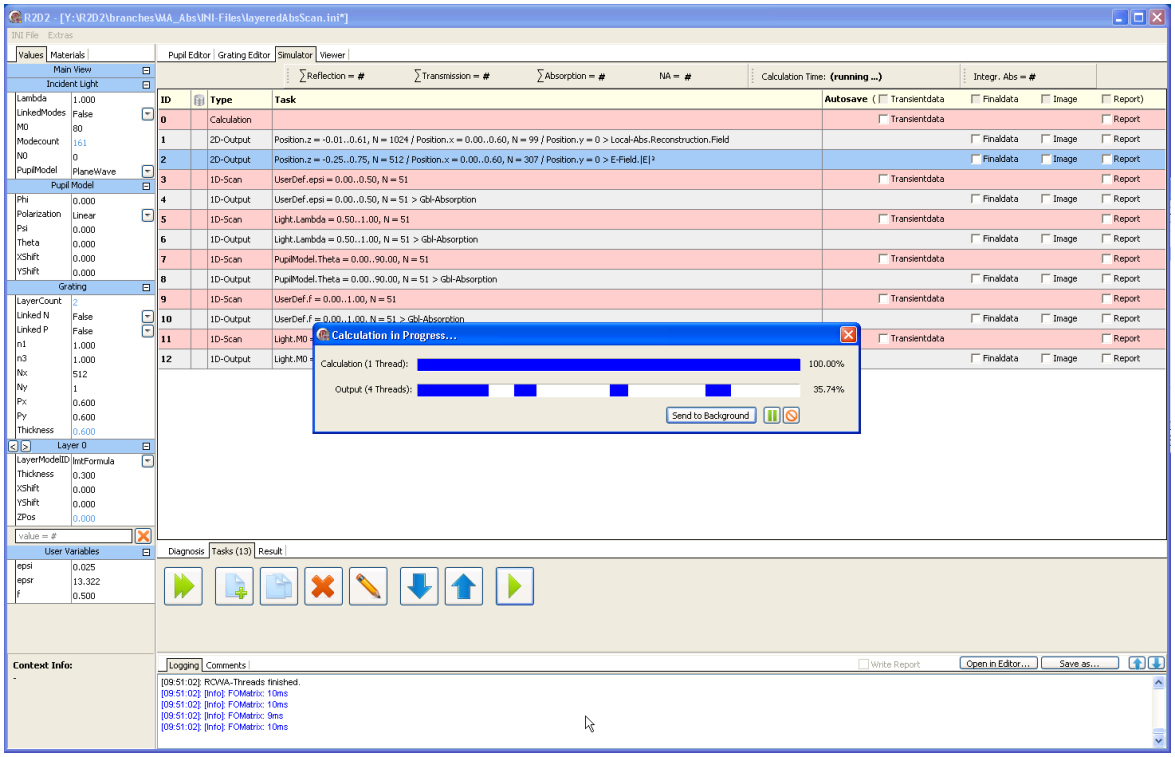


Figure B.3: Screenshot of the `Simulator`. The background shows the task list. In the foreground in the background during a calculation

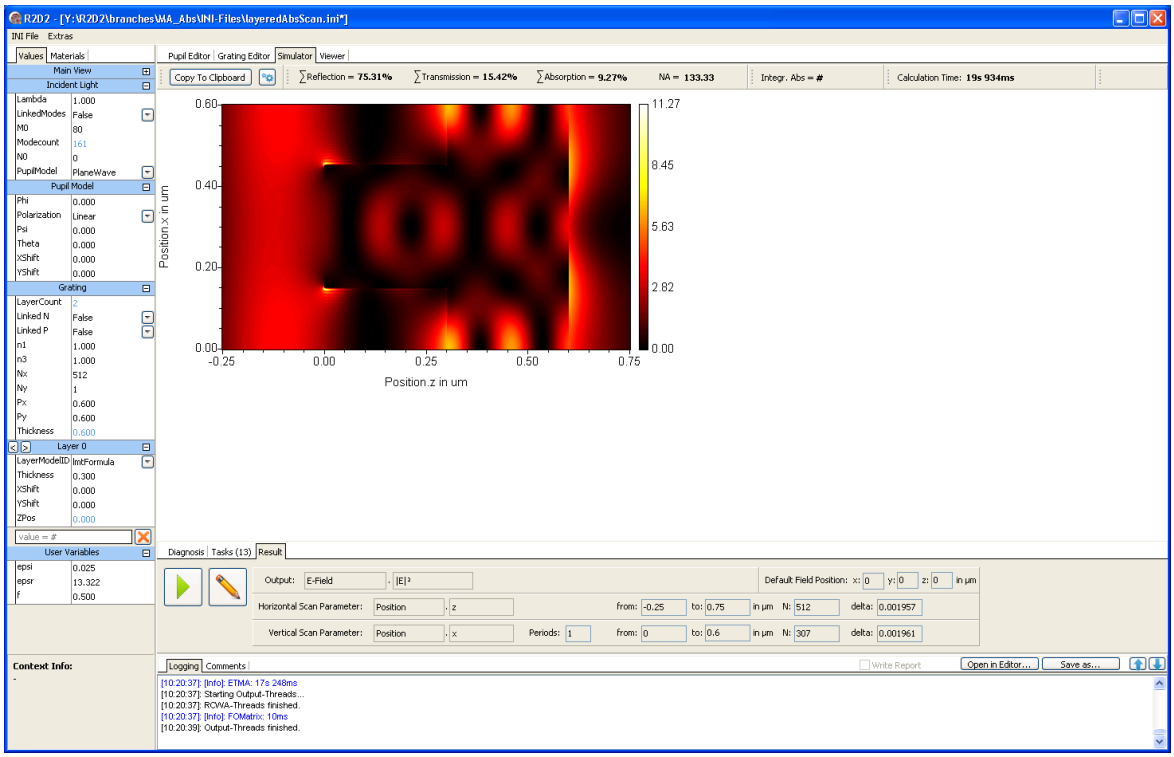


Figure B.4: Screenshot of the `Viewer`, showing the result of a field calculation and the corresponding task at the bottom

B.0.5 Outlook

This section provided a brief description of the simulation software that was developed during the time of this work and used to perform all RCWA related simulations presented here. The software already features multithreading for parameter scans and OutputTasks. However, there are also algorithms for the parallel calculation of the eigenvalue problem, which is the most time-critical element of a single RCWA calculation. This would improve the simulation speed for high truncation limits. Besides, the multithreading capabilities could also be extended to GPUs and distributed computer networks in the future.

Bibliography

[1] N. van der Aa. "Diffraction grating theory with RCWA or the C method". In: *Progress in Industrial Mathematics at ECMI 2004*. Springer, 2006, 99–103.

[2] D. Aspnes et al. "Investigation of effective-medium models of microscopic surface roughness by spectroscopic ellipsometry". In: *Physical Review B* 20.8 (1979), p. 3292.

[3] M. Auer. "Experimental setup for the measurement of resonant effects in high-contrast gratings". In: *Annual Report 2013 of Chair of Optoelectronics, Heidelberg University* (2013).

[4] M. Auer. "Rigorous Light Propagation in Waveguides as one application of the Localized Input Field-RCWA (LIF-RCWA)". In: *Annual Report 2013 of Chair of Optoelectronics, Heidelberg University* (2014).

[5] M. Auer and K.-H. Brenner. "Enhancement of photodetector responsivity in standard SOI CMOS processes by introducing resonant grating structures". In: *JEOS:RP* 6 (2011).

[6] M. Auer and K.-H. Brenner. "Localized input fields in rigorous coupled-wave analysis". In: *JOSA A* 31.11 (2014), pp. 2385–2393.

[7] K. Aydin et al. "Broadband polarization-independent resonant light absorption using ultrathin plasmonic super absorbers". In: *Nature communications* 2 (2011), p. 517.

[8] M. Azadeh. "Fiber Optics Engineering". In: Optical Networks. Springer US, 2009. Chap. Attenuation, p. 136.

[9] D. Basting and G. Marowsky. "Excimer Laser Technology". In: Advanced texts in physics. Springer, 2005. Chap. Challenges of EUV Lithography, 401 ff..

[10] N. Bonod et al. "Light transmission through a subwavelength microstructured aperture: electromagnetic theory and applications". In: *Optics communications* 245.1 (2005), pp. 355–361.

[11] M. Born et al. *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*. Cambridge University Press, 1999.

[12] R. Bräuer and O. Bryngdahl. "Electromagnetic diffraction analysis of two-dimensional gratings". In: *Optics communications* 100.1 (1993), pp. 1–5.

[13] K.-H. Brenner. "Aspects for calculating local absorption with the rigorous coupled-wave method". In: *Optics express* 18.10 (2010), pp. 10369–10376.

[14] K.-H. Brenner. *Skript zur Vorlesung 'Computational Optics'*. University of Heidelberg, 2012/2013.

[15] K.-H. Brenner. *Skript zur Vorlesung 'Grundlagen der Optik'*. University of Heidelberg, 2007.

[16] K.-H. Brenner and M. Auer. "Lokalität von Quellen und Senken in der exakten optischen Simulation". In: *DGAO 2014 in Karlsruhe*. 2014.

[17] K.-H. Brenner and W. Singer. “Light propagation through microlenses: a new simulation method”. In: *Applied optics* 32.26 (1993), pp. 4984–4988.

[18] C. Burckhardt. “Diffraction of a plane wave at a sinusoidally stratified dielectric grating”. In: *JOSA* 56.11 (1966), pp. 1502–1508.

[19] M. Cadilhac. “Some Mathematical Aspects of the Grating Theory”. In: *Electromagnetic Theory of Gratings*. Ed. by R. Petit. Vol. 22. Topics in Current Physics. Springer Berlin Heidelberg, 1980, pp. 53–62.

[20] J. Chandezon et al. “A new theoretical method for diffraction gratings and its numerical application”. In: *Journal of Optics* 11.4 (1980), p. 235.

[21] J. Chilwell and I. Hodgkinson. “Thin-films field-transfer matrix theory of planar multilayer waveguides and reflection from prism-loaded waveguides”. In: *J. Opt. Soc. Am. A* 1.7 (July 1984), pp. 742–753.

[22] E. Chu. “Discrete and Continuous Fourier Transforms: Analysis, Applications and Fast Algorithms”. In: CRC Press, 2008. Chap. 3.8, 61 ff.

[23] R.-S. Chu and T. Tamir. “Bragg diffraction of Gaussian beams by periodically modulated media”. In: *JOSA* 66.3 (1976), pp. 220–226.

[24] R. Dalven. “Empirical relation between energy gap and lattice constant in cubic semiconductors”. In: *Physical Review B* 8.12 (1973), p. 6033.

[25] C. G. Darwin. “LXXVIII. The theory of X-ray reflexion. Part II”. In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 27.160 (1914), pp. 675–690.

[26] M. P. Davidson. “Analytic waveguide solutions and the coherence probe microscope”. In: *Microelectronic Engineering* 13.1–4 (1991), pp. 523–526.

[27] T. Delort and D. Maystre. “Finite-element method for gratings”. In: *J. Opt. Soc. Am. A* 10.12 (Dec. 1993), pp. 2592–2601.

[28] H. DeVoe. “Optical properties of molecular aggregates. I. Classical model of electronic absorption and refraction”. In: *The Journal of chemical physics* 41.2 (1964), 393–400.

[29] H. DeVoe. “Optical properties of molecular aggregates. II. Classical theory of the refraction, absorption, and optical activity of solutions and crystals”. In: *The Journal of chemical physics* 43.9 (1965), pp. 3199–3208.

[30] B. T. Draine and P. J. Flatau. “Discrete-Dipole Approximation For Scattering Calculations”. In: *J. Opt. Soc. Am. A* 11.4 (Apr. 1994), 1491–1499.

[31] B. T. Draine and P. J. Flatau. “User guide for the discrete dipole approximation code DDSCAT 7.3”. In: *arXiv preprint arXiv:1305.6497* (2013).

[32] K. Edee et al. “Perturbation method for the rigorous coupled wave analysis of grating diffraction”. In: *Optics express* 18.25 (2010), pp. 26274–26284.

[33] G. Even and M. Medina. “Digital Logic Design: A Rigorous Approach”. In: Cambridge University Press, 2012, p. 133.

[34] S. Fairchild. “CMOS, The Ideal Logic Family”. In: *AN-77, Jan* (1983), p. 133.

[35] M. Feit and J. Fleck. “Light propagation in graded-index optical fibers”. In: *Applied optics* 17.24 (1978), pp. 3990–3998.

[36] M. Fertig and K.-H. Brenner. “Vector wave propagation method”. In: *JOSA A* 27.4 (2010), pp. 709–717.

[37] D. G. Flagello et al. “Theory of high-NA imaging in homogeneous thin films”. In: *JOSA A* 13.1 (1996), pp. 53–64.

[38] P. J. Flatau and B. T. Draine. “Fast near field calculations in the discrete dipole approximation for regular rectilinear grids”. In: *Opt. Express* 20.2 (Jan. 2012), 1247–1252.

[39] J. W. Gibbs. “On multiple algebra”. In: *Proceedings of the American Association for the Advancement of Science* 35 (1886), pp. 37–66.

[40] P. Götz et al. “Normal vector method for the RCWA with automated vector field generation”. In: *Optics express* 16.22 (2008), pp. 17295–17301.

[41] G. Granet and B. Guizal. “Efficient implementation of the coupled-wave method for metallic lamellar gratings in TM polarization”. In: *JOSA A* 13.5 (1996), 1019–1023.

[42] R. M. Gray. “Toeplitz and Circulant Matrices”. In: Now Publishers, 2006. Chap. 3 & 4.

[43] C. Hafner and R. Ballisti. “The multiple multipole method (MMP)”. In: *COMPEL - The international journal for computation and mathematics in electrical and electronic engineering* 2.1 (1983), pp. 1–7.

[44] C. Hafner. *The generalized multipole technique for computational electromagnetics*. Artech House on Demand, 1990.

[45] F. J. Harris. “On the use of windows for harmonic analysis with the discrete Fourier transform”. In: *Proceedings of the IEEE* 66.1 (1978), pp. 51–83.

[46] O. Heaviside. “The Induction of Currents in Cores”. In: *The Electrician* 13 (1884), pp. 133–4.

[47] C. Heine and R. H. Morf. “Submicrometer gratings for solar energy applications”. In: *Applied Optics* 34.14 (1995), pp. 2476–2482.

[48] J. J. Hench and Z. STRAKOŠ. “The RCWA method-a case study with open questions and perspectives of algebraic computations”. In: *Electronic Transactions on Numerical Analysis* 31 (2008), pp. 331–357.

[49] H. Hertz. “Ueber die Grundgleichungen der Elektrodynamik für bewegte Körper”. ger. In: *Nachrichten von der Königl. Gesellschaft der Wissenschaften und der Georg-Augusts-Universität zu Göttingen* 41 (1890), pp. 369–399.

[50] H. Hertz. “Ueber die Grundgleichungen der Elektrodynamik für ruhende Körper”. ger. In: *Nachrichten von der Königl. Gesellschaft der Wissenschaften und der Georg-Augusts-Universität zu Göttingen* 41 (1890), pp. 577–624.

[51] A. P. Hibbins. “Grating coupling of surface plasmon polaritons at visible and microwave frequencies”. PhD thesis. University of Exeter, 2000.

[52] J. Hugonin et al. “Plane-wave expansions used to describe the field diffracted by a grating”. In: *JOSA* 71.5 (1981), pp. 593–598.

[53] H. Ichikawa. “Electromagnetic analysis of diffraction gratings by the finite-difference time-domain method”. In: *J. Opt. Soc. Am. A* 15.1 (Jan. 1998), pp. 152–157.

[54] K. Inoue and K. Ohtaka. *Photonic Crystals: Physics, Fabrication and Applications*. Springer Series in Optical Sciences. Springer Berlin Heidelberg, 2013.

[55] J. D. Jackson. “Classical Electrodynamics”. In: John Wiley & Sons, 1999. Chap. 7.10, pp. 333–335.

[56] X. Jing and Y. Jin. “Transmittance analysis of diffraction phase grating”. In: *Applied Optics* 50.9 (2011), pp. C11–C18.

[57] C. Johnson. *Numerical Solution of Partial Differential Equations by the Finite Element Method*. Dover Books on Mathematics Series. Dover Publications, Incorporated, 2012.

[58] C. Johnson and L. Pedersen. “Problems and Solutions in Quantum Chemistry and Physics”. In: Dover Books on Chemistry Series. Dover, 1974. Chap. 2.3.

[59] F. G. Kaspar. “Diffraction by thick, periodically stratified gratings with complex dielectric constant”. In: *JOSA* 63.1 (1973), pp. 37–45.

[60] N. Kerwien. “Zum Einfluss von Polarisationseffekten in der mikroskopischen Bildentstehung”. PhD thesis. Institut für technische Optik, Universität Stuttgart, 2007.

[61] N. Kerwien et al. “Vectorial thin-element approximation: a semirigorous determination of Kirchhoff’s boundary conditions”. In: *JOSA A* 24.4 (2007), pp. 1074–1084.

[62] A. Khavasi et al. “Three-dimensional diffraction analysis of gratings based on Legendre expansion of electromagnetic fields”. In: *JOSA B* 24.10 (2007), pp. 2676–2685.

[63] H. Kim and B. Lee. “Pseudo-Fourier modal analysis of two-dimensional arbitrarily shaped grating structures”. In: *JOSA A* 25.1 (2008), pp. 40–54.

[64] B. H. Kleemann. “Elektromagnetische Analyse von Oberflächengittern von IR bis XUV mittels einer parametrisierten Randintegralmethode: Theorie, Vergleich und Anwendungen”. PhD thesis. 2003.

[65] B. H. Kleemann et al. “Integral equation method with parametrization of grating profile theory and experiments”. In: *journal of modern optics* 43.7 (1996), 1323–1349.

[66] H. Kogelnik. “Coupled wave theory for thick hologram gratings”. In: *Bell System Technical Journal* 48.9 (1969), pp. 2909–2947.

[67] P. Lalanne and M. P. Jurek. “Computation of the near-field pattern with the coupled-wave method for transverse magnetic polarization”. In: *Journal of modern optics* 45.7 (1998), pp. 1357–1374.

[68] P. Lalanne and G. M. Morris. “Highly improved convergence of the coupled-wave method for TM polarization”. In: *JOSA A* 13.4 (1996), pp. 779–784.

[69] P. Lalanne et al. “Numerical analysis of a slit-groove diffraction problem”. In: *Journal of the European Optical Society-Rapid Publications* 2 (2007).

[70] P. Langois and R. Beaulieu. “Phase relief gratings with conic section profile used in the production of multiple beams”. In: *Appl. Opt.* 29.23 (Aug. 1990), pp. 3434–3439.

[71] U. Levy et al. “Thin element approximation for the analysis of blazed gratings: simplified model and validity limits”. In: *Optics communications* 229.1 (2004), 11–21.

[72] L. Li. “Bremmer series, R-matrix propagation algorithm, and numerical modeling of diffraction gratings”. In: *JOSA A* 11.11 (1994), pp. 2829–2836.

[73] L. Li. “Formulation and comparison of two recursive matrix algorithms for modeling layered diffraction gratings”. In: *JOSA A* 13.5 (1996), 1024–1035.

[74] L. Li. “Fourier modal method for crossed anisotropic gratings with arbitrary permittivity and permeability tensors”. In: *Journal of Optics A: Pure and Applied Optics* 5.4 (2003), p. 345.

[75] L. Li. “Multilayer modal method for diffraction gratings of arbitrary profile, depth, and permittivity”. In: *JOSA A* 10.12 (1993), pp. 2581–2591.

[76] L. Li. “New formulation of the Fourier modal method for crossed surface-relief gratings”. In: *JOSA A* 14.10 (1997), pp. 2758–2767.

[77] L. Li. “Use of Fourier series in the analysis of discontinuous periodic structures”. In: *JOSA A* 13.9 (1996), pp. 1870–1876.

[78] N. Liu et al. “Three-dimensional photonic metamaterials at optical frequencies”. In: *Nature materials* 7.1 (2008), pp. 31–37.

[79] E. Loewen and E. Popov. *Diffraction Gratings and Applications*. Optical Science and Engineering. Taylor & Francis, 1997.

[80] M. Mansuripur. “Distribution of light at and near the focus of high-numerical-aperture objectives”. In: *JOSA A* 3.12 (1986), pp. 2086–2093.

[81] J. C. Maxwell. “Xxv. on physical lines of force: Part i.–the theory of molecular vortices applied to magnetic phenomena”. In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 21.139 (1861), 161–175.

[82] D. Maystre. “Integral Methods”. In: *Electromagnetic Theory of Gratings*. Ed. by R. Petit. Vol. 22. Topics in Current Physics. Springer Berlin Heidelberg, 1980.

[83] D. Maystre and E. Popov. “Integral method for gratings”. In: *arXiv preprint arXiv: 1406.0313* (2014).

[84] D. Melville and R. Blaikie. “Super-resolution imaging through a planar silver layer”. In: *Optics Express* 13.6 (2005), pp. 2127–2134.

[85] G. Mie. “Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen”. In: *Annalen der Physik* 330.3 (1908), pp. 377–445.

[86] R. Millar. “On the Rayleigh assumption in scattering by a periodic surface”. In: *Mathematical Proceedings of the Cambridge Philosophical Society*. Vol. 65. 03. Cambridge Univ Press. 1969, pp. 773–791.

[87] M. Moharam and T. Gaylord. “Diffraction analysis of dielectric surface-relief gratings”. In: *JOSA* 72.10 (1982), pp. 1385–1392.

[88] M. Moharam and T. Gaylord. “Rigorous coupled-wave analysis of planar-grating diffraction”. In: *JOSA* 71.7 (1981), pp. 811–818.

[89] M. Moharam and T. Gaylord. “Three-dimensional vector coupled-wave analysis of planar-grating diffraction”. In: *JOSA* 73.9 (1983), pp. 1105–1112.

[90] M. Moharam et al. “Bragg diffraction of finite beams by thick gratings”. In: *JOSA* 70.3 (1980), pp. 300–304.

[91] M. Moharam et al. “Formulation for stable and efficient implementation of the rigorous coupled-wave analysis of binary gratings”. In: *JOSA A* 12.5 (1995), pp. 1068–1076.

[92] M. Moharam et al. “Stable implementation of the rigorous coupled-wave analysis for surface-relief gratings: enhanced transmittance matrix approach”. In: *JOSA A* 12.5 (1995), pp. 1077–1086.

[93] N. Moll et al. “Polarization-Independent Photodetectors With Enhanced Responsivity in a Standard Silicon-on-Insulator Complementary Metal–Oxide–Semiconductor Process”. In: *Journal of Lightwave Technology* 27.21 (2009), pp. 4892–4896.

[94] E. Moreno et al. “Multiple multipole method with automatic multipole setting applied to the simulation of surface plasmons in metallic nanostructures”. In: *JOSA A* 19.1 (2002), pp. 101–111.

[95] P. Müllner. “Fundamental characteristics of the soi slot waveguide structure”. PhD thesis. University of Vienna, 2010.

[96] H. Nalwa. “Photodetectors and Fiber Optics”. In: Elsevier Science, 2012. Chap. 2.2.

[97] M. Nevière and E. Popov. “Light propagation in periodic media: differential theory and design”. In: CRC Press, 2002. Chap. 1.1.3, 1.2.2, 1.2.3 - 5.

[98] M. Nevière et al. “Theory of conducting gratings and their applications to Optics”. In: *Nouvelle Revue d’Optique* 5.2 (1974), p. 65.

[99] M. Notomi. “Theory of light propagation in strongly modulated photonic crystals: Refractionlike behavior in the vicinity of the photonic band gap”. In: *Physical Review B* 62.16 (2000), p. 10696.

[100] J. B. Pendry. “Negative refraction makes a perfect lens”. In: *Physical review letters* 85.18 (2000), p. 3966.

[101] S. Peng et al. “Theory of periodic dielect waveguides”. In: *Microwave Theory and Techniques, IEEE Transactions on* 23.1 (1975), pp. 123–133.

[102] R. Petit. “Etude théorique de la diffraction d’une onde plane et monochromatique par un réseau métallique infiniment conducteur”. In: *Appl. Opt.* 4.12 (Dec. 1965), pp. 1551–1554.

[103] R. Petit and L. Botten. *Electromagnetic theory of gratings*. Topics in current physics. Springer-Verlag, 1980.

[104] D. A. Pommet et al. “Limits of scalar diffraction theory for diffractive phase elements”. In: *JOSA A* 11.6 (1994), pp. 1827–1834.

[105] E. Popov. “Gratings: Theory and Numeric Applications, Second Revisited Edition”. In: AMU (PUP), 2014. Chap. 7 Differential method for periodic structures.

[106] E. Popov. *Gratings: theory and numeric applications*. Popov, Institut Fresnel, 2012.

[107] E. Popov and M. Nevière. “Maxwell equations in Fourier space: fast-converging formulation for diffraction by arbitrary shaped, periodic, anisotropic media”. In: *JOSA A* 18.11 (2001), pp. 2886–2894.

[108] D. Rittenhouse. “An optical problem proposed by F. Hopkinson and solved”. In: *J. Am. Phil. Soc* 201 (1786), pp. 202–206.

[109] P. S. J. Russell. “Power conservation and field structures in uniform dielectric gratings”. In: *JOSA A* 1.3 (1984), pp. 293–299.

[110] B. E. A. Saleh and M. C. Teich. “"Planar dielectric waveguides" in Fundamentals of Optics”. In: Wiley, 2007, pp. 299–305.

[111] H. Schmid et al. “Template-assisted selective epitaxy of III–V nanoscale devices for co-planar heterogeneous integration with Si”. In: *Applied Physics Letters* 106.23 (2015), p. 233101.

[112] T. Schuster. “Simulation von Lichtbeugung an Kreuzgitter-Strukturen und deren Anwendung in der Scatterometrie”. PhD thesis. Institut fuer Technische Optik der Universitaet Stuttgart, 2010.

[113] T. Schuster et al. “Normal vector method for convergence improvement using the RCWA for crossed gratings”. In: *JOSA A* 24.9 (2007), pp. 2880–2890.

[114] E. Silberstein et al. “Use of grating theories in integrated optics”. In: *JOSA A* 18.11 (2001), pp. 2865–2875.

[115] H. Singer. “Numerische Feldberechnung mit Hilfe von Multipolen”. German. In: *Archiv für Elektrotechnik* 59.4 (1977), pp. 191–195.

[116] L. Solymar. “A general two-dimensional theory for volume holograms”. In: *Applied Physics Letters* 31.12 (1977), pp. 820–822.

[117] A. Sommerfeld. “Mathematische Theorie der Diffraction”. German. In: *Mathematische Annalen* 47.2-3 (1896), pp. 317–374.

[118] R. Soref. “The Achievements and Challenges of Silicon Photonics”. In: *Advances in Optical Technologies* 2008.472305 (2008).

[119] B. Stout et al. “Mie scattering by an anisotropic object. Part II. Arbitrary-shaped object: differential theory”. In: *JOSA A* 23.5 (2006), pp. 1124–1134.

[120] J. Suratteau et al. “Sur la détermination numérique des efficacités de certains réseaux diélectriques profonds”. In: *Journal of optics* 14.6 (1983), pp. 273–288.

[121] S. Sze and K. Ng. “Physics of Semiconductor Devices”. In: Wiley, 2006. Chap. Appendix G, p. 790.

[122] D. Taillaert et al. “Grating couplers for coupling between optical fibers and nanophotonic waveguides”. In: *Japanese Journal of Applied Physics* 45.8R (2006), p. 6071.

[123] E. L. Tan. “Note on formulation of the enhanced scattering- (transmittance-) matrix approach”. In: *JOSA A* 19.6 (2002), pp. 1157–1161.

[124] G. Tayeb. “The method of fictitious sources applied to diffraction gratings”. In: *Applied Computational Electromagnetics Society Journal* 9.3 (1994), 90–100.

[125] A. Tishchenko. “Generalized source method: New possibilities for waveguide and grating problems”. In: *Optical and Quantum Electronics* 32.6-8 (2000), pp. 971–980.

[126] B. Trauter. “Untersuchungen zum inversen, elektromagnetischen Gitterbeugungsproblem mit Anwendungen in der Gittercharakterisierung”. PhD thesis. Universität Mannheim, 2009.

[127] T. Vallius. “Comparing the Fourier modal method with the C method: analysis of conducting multilevel gratings in TM polarization”. In: *J. Opt. Soc. Am. A* 19.8 (Aug. 2002), pp. 1555–1562.

[128] V. Vemlago. “The electrodynamics of substances with simultaneously negative values of ϵ and μ”. In: *Sov Phys Usp* 10.4 (1968), pp. 509–514.

[129] Y. A. Vlasov et al. “On-chip natural assembly of silicon photonic bandgap crystals”. In: *Nature* 414.6861 (2001), pp. 289–293.

[130] S. Wang and R. Magnusson. “Theory and applications of guided-mode resonance filters”. In: *Applied optics* 32.14 (1993), pp. 2606–2613.

[131] J. Wauer and T. Rother. “Considerations to Rayleigh’s hypothesis”. In: *Optics Communications* 282.3 (2009), pp. 339–350.

[132] C. Webb and J. Jones. “Handbook of Laser Technology and Applications: Principles”. In: Handbook of Laser Technology and Applications. Institute of Physics, 2004. Chap. A2.1.10, p. 111.

[133] S.-D. Wu et al. “Three-dimensional converging-diverging Gaussian beam diffraction by a volume grating”. In: *JOSA A* 22.7 (2005), pp. 1293–1303.

[134] K. S. Yee et al. “Numerical solution of initial boundary value problems involving Maxwell’s equations in isotropic media”. In: *IEEE Trans. Antennas Propag* 14.3 (1966), pp. 302–307.

[135] W.-L. Yeh and Y.-P. Chiou. “A stable approach to conical diffraction of nearly lossless metallic gratings”. In: *Optical and Quantum Electronics* 47.3 (2015), pp. 535–543.

[136] T. Yoshizawa. “Handbook of Optical Metrology: Principles and Applications”. In: CRC Press, 2009. Chap. Incadescent Lamps and Tungsten Haogen Lamps.

[137] T. Young. “On the theory of light and colors”. In: *Phil. Trans. II* (1803), pp. 399–408.

[138] H. Zappe. “Fundamentals of Micro-Optics”. In: Cambridge University Press, 2010. Chap. 2.6.2, p. 47.

Acknowledgement

First and foremost, I express my gratitude to my advisor, Professor Dr. Brenner, head of the chair of optoelectronics at the Institute of Computer Engineering of the University of Heidelberg (ZITI), for his constant support, motivation and numerous inspiring discussions. I especially appreciated his open door policy, which he always maintained despite of all his other obligations.

I also thank my second reviewer Professor Dr. Osten, director of the Institute for Applied Optics at the University of Stuttgart (ITO), for his time and interest in my research.

Special mention is due to the colleagues from IBM: Dr. Matthias Fertig, Dr. Nikolaj Moll, Dr. Thomas Morf, Dr. Thilo Stöferle and Dr. Jens Hofrichter, who provided a substantial contribution to the development of the photodetector.

I am also thankful to my fellows at the chair of optoelectronics: Dr. Xiyuan Liu, Andre Junker and Tim Stenau for their valuable inputs in theoretical discussions, Wolfgang Stumpfs for his support in technical questions and his dear friendship and Sabine Volk for sharing her great organizational talent on numerous occasions.

Last but not least, my thanks go to my family and my friends: I thank my parents as well as my brothers for being a true and reliable constant in my life and all my friends for their patience and unquestioned support over the last years. A special mention goes to Stanislav Jakuschevskij, Martin Kolb, Steffen Lohaus, Arthur Müller, Dmitry Poborchiy, Kiril Schewzow and Martin Schlichte. Finally, I thank my beloved fiancée Sabrina Zwykiel for her love and confidence in me, which gave me the strength to complete this work.