Convergence Study of the Fourier Modal Method for Nano-optical Scattering Problems in Comparison with the Finite Element Method

Diplomarbeit

vorgelegt von

Philipp Gutsche

Betreuer

Prof. Dr. Thomas Judd
Arbeitsgruppe Computational Quantenphysik
Physikalisches Institut
Mathematisch-Naturwissenschaftliche Fakultät
Eberhard Karls Universität Tübingen

und

Prof. Dr. Frank Schmidt
Arbeitsgruppe Mathematische Nano-Optik
Abteilung Numerische Analysis und Modellierung
Bereich Numerische Mathematik
Konrad-Zuse-Zentrum für Informationstechnik Berlin

November, 2014

Eidesstattliche Erklärung

Hiermit wird bestätigt, dass die Diplomarbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt wurden.

_____

Datum, Unterschrift

Contents

Abbreviations

ASR	Adaptive Spatial Resolution
CD	Circular Dichroism
CoDo	Computational Domain
CPL	Circularly Polarized Light
EET	Eigenmode Expansion Technique
EME	Eigenmode Expansion
EUV	Extreme Ultraviolet Lithography
FDTD	Finite-Difference Time-Domain
FEM	Finite Element Method
FFF	Fast Fourier Factorization
FFR	Fourier Factorization Rules
FFT	Fast Fourier Transfom
FMM	Fourier Modal Method
FT	Fourier Transform
PhC	Photonic Crystal
PML	Perfectly Matched Layers
PWE	Plane Wave Expansion
RCWA	Rigorous Coupled Wave Analysis
S^4	Stanford Stratified Structure Solver
SPS	Single-Photon Source
TE	Transverse Electric
TM	Transverse Magnetic

Abstract

Nano-optische Streuprobleme spielen eine wichtige Rolle in unserer modernen, technologischen Gesellschaft. Computer, Smartphones und unzählige elektronische Geräte werden von der Halbleiterindustrie hergestellt. Hierfür werden sowohl Fotomasken als auch die optische Prozesskontrolle verwendet. Auch die digitale Welt, z.B. das Internet, basiert auf optischer Datenübertragung und die nächste Generation der Computer sind vermutlich so genannte Quantencomputer, die optische Phänomene nutzen. Weiterhin führt der globale wirtschaftliche Aufschwung zu einem erhöhten Energiebedarf, der mit nachhaltigen Energieformen wie der Nutzung der Sonneneinstrahlung gedeckt werden kann. Außerdem entstehen Innovationen in den Ingenieurwissenschaften aus dem Verständnis fundamentaler physikalischer Vorgänge, wie z.B. den optischen Eigenschaften von unsymmetrischen, so genannten chiralen, Strukturen.

Um diese optischen Prozesse zu verstehen hat sich in der Physik ein weitverbreitetes Modell etabliert: die so genannten Maxwell-Gleichungen. Sie wurden 1862 von James Clerk Maxwell formuliert und beschreiben die Wechselwirkungen von Licht und Materie. Zur Lösung dieser Gleichungen für komplizierte realistische Probleme reicht einfache analytische Mathematik nicht aus. Vielmehr werden hierfür Computer-Simulationen eingesetzt für die eine große Zahl verschiedener numerischer Methoden zur Verfügung steht. Das Gebiet der Numerik befasst sich mit der Fragestellung, für welches Problem welche Methode am besten geeignet ist. Vereinfacht kann hier zwischen langer Rechenzeit für so genannte Zeitverfahren (z.B. Finite-Differenzen-Methode) und hohem Speicherbedarf so genannter Frequenzbereich-Verfahren (z.B. Fourier-Moden-Methode und Finite-Elemente-Methode) unterschieden werden.

Das Ziel dieser Arbeit ist die Untersuchung der Anwendbarkeit der Fourier-Moden-Methode (FMM, Fourier Modal Method) für nano-optische Streuprobleme. Da wie bereits erwähnt generell keine einfachen analytischen Lösungen für moderne Fragestellungen dieser Art existieren, wird in der vorliegenden Arbeit die Finite-Elemente-Methode (FEM, Finite Element Method) verwendet, um das Verhalten der FMM zu überprüfen. Für die FEM existiert eine weit entwickelte mathematische Konvergenz-Theorie, die es ermöglicht den Fehler der Ergebnisse dieses Verfahrens abzuschätzen und zu kontrollieren. Im Gegensatz dazu ist es bisher nicht möglich die Approximations-Eigenschaften der FMM rigoros zu behandeln. Deshalb kann nicht sichergestellt werden, dass diese Methode für einen erhöhten numerischen Aufwand grundsätzlich bessere Ergebnisse liefert. Daher bleibt die Frage, ob dieses numerische Verfahren für sämtliche Problemstellungen konvergiert, unbeantwortet.

Trotz dieser Unsicherheit ist die FMM ein etabliertes Berechnungsverfahren, dessen Ursprünge bis in die 1980er Jahre reichen. Sie wurde unlängst zur Optimierung von Solarzellen [19] und zur Studie der optischen Eigenschaften von so genannten Einzel-Photonen-Quellen [41] benutzt. Einzel-Photonen-Quellen sind ein wichtiges Werkzeug der Quantenkryptographie, die die digitale Sicherheit und Verschlüsselungstechniken revolutionieren könnte. Darüber hinaus wurde das Problem der winkelabhängigen Filterung elektromagnetischer Wellen mit Hilfe der FMM teilweise gelöst [77]: Dazu wurde ein Spiegel entwickelt, der unter einem bestimmten Betrachtungswinkel transparent wird. Außerdem wurde basierend auf der Fourier-Moden-Methode eine verbesserte Methode zur Entwicklung von so genannten Photonischen-Kristall-Wellenleitern entwickelt [15]. Photonische Kristalle werden beispielsweise für optisch-biologische Sensoren und zur Entwicklung neuer Halbleiter-Schaltungen genutzt. Des Weiteren gibt es Ansätze, die die Vorteile der FMM und der FEM kombinieren, und somit die Möglichkeit bieten schnell genaue Simulationsergebnisse zu erhalten. Diese Ansätze basieren auf ähnlichen Ansätzen wie so genannte Gebietsverteilungs-Verfahren [88]. Die Kopplung mehrerer Gebiete kann hier mit dem Streumatrix-Formalismus erreicht werden, der auch in der FMM verwendet wird.

Die vorliegende Arbeit beginnt mit der Darstellung der Maxwell-Gleichungen im Besonderen für periodische Probleme, da die FMM inhärent periodisch ist. Sie entstand als Teil der Theorie von Beugungsgittern, weshalb die Grundlagen dieser Theorie knapp dargestellt werden. Darüber hinaus werden zwei nano-optische Phänomene erläutert, die so genannte optische Chiralität asymmetrischer

Partikel und die elektronisch-optische Kopplung von Linsen für Einzel-Photonen-Quellen. Zum Hintergrund dieser Arbeit zählen ebenfalls die Motivierung und die Definition der Fehlermaße, die im weiteren Verlauf wesentlich sind.

Die FMM ist unter vielen Synonymen bekannt und deren erste Formulierung wurde mit der Abkürzung RCWA benannt [54]. Der moderne Begriff, Fourier-Moden-Methode, wurde durch Li geprägt, der ebenfalls den mathematischen Hintergrund der so genannten inversen Regel entwickelte [43]. Dieses Verfahren zur Faltung im Fourier-Raum bedeutete den Durchbruch der FMM, da die Konvergenz für so genannte metallische TM Beugungsgitter drastisch beschleunigt wurde. Deshalb wird dessen Beweis [1] zusammengefasst und erwähnt, dass es bisher keine physikalische Motivation für dieses Verfahren gibt. Im Anschluss werden die bekannten Varianten der FMM [49] rekapituliert und vorgestellt: Im Laufe der Entwicklung der FMM wurde die korrekte Verwendung der so genannten Fourier-Faktorisierungs-Regeln (Fourier Factorization Rules) von zwei- auf drei dimensionale Probleme erweitert. Hierfür wurden Normalen-Vektorfelder [75] bzw. eine Jones Polarisationsbasis [2] verwendet. Diese Methoden werden auf ihre Konvergenz untersucht. Das neuere Konzept der räumlich adaptiven Auflösung (Adaptive Spatial Resolution) der FMM [18] ist hingegen nicht Teil dieser Konvergenzstudie.

Die Grundlagen der Finiten-Elemente-Methode [57] werden ebenso dargestellt wie die Aufteilung in ein Innenraum- und ein Außenraum-Problem, für die die so genannte schwache Formulierung der Maxwell-Gleichungen hergeleitet wird. Fortgeschrittene Konzepte zur Behandlung des Außenraum-Problems werden knapp motiviert und eine Erweiterung der FEM, die *hp*-Adaptivität [16], anhand eines numerischen Beispiels vorgestellt. Des Weiteren werden der Ablauf eines Algorithmus zur Berechnung von 2D-Schnitten aus einem dreidimensionalen Finite-Elemente-Gitter erläutert und dessen Ergebnisse an einem Beispiel illustriert.

Sämtliche Simulationen dieser Arbeit wurden für Ergebnisse der FMM mit dem frei verfügbaren Programm S^4 [70] und im Fall der FEM mit dem Paket *JCMsuite* [32] berechnet. Hiermit werden die Eigenschaften der Glättung der Permittivität innerhalb der FMM und die so genannte schnelle Fourier Transformation untersucht. In beiden Fällen ergibt sich keine Verbesserung der FMM gegenüber geschlossener Integrationsformeln für die Fourier Transformation der Materialparameter. Das so genannte Gibbs Phänomen wird als bekanntes Problem der Fourier Transformation diskontinuierlicher Funktionen näher analysiert. Dazu wird die FMM mit der analytischen Fourier Transformation der Permittivität verglichen. Die Ergebnisse des zweiten Modells werden mit Hilfe der FEM berechnet und es zeigt sich, dass die ebenen Wellen-Basis der FMM einen signifikanten Fehler gegenüber der reinen Approximation der Materialien mit sich bringt.

Obwohl die Vollständigkeit der ebenen Wellen Basis für komplexe Permittivitäten nicht bewiesen ist [39], wird dessen Basiseigenschaft üblicherweise vorausgesetzt. Deshalb wird eine Fotomaske, die im Bereich extrem ultravioletter Strahlung eingesetzt wird, untersucht. Diese beinhaltet metallische Streuer und die vorliegenden Ergebnisse zeigen, dass Fehler in dieser Simulation vor allem an Schicht-Übergängen und in metallischen Strukturen entstehen. Dies trifft auch auf den so genannten Stufen-Effekt [55] zu, der für einen zweidimensionalen photonischen Kristall untersucht wird. Hier wird deutlich, dass die Einführung einer hohen Zahl an Schichten durch diese Methode einen Fehler in der Phase der Fourier Koeffizienten nach sich zieht. Der Fehler des Betrags der Fourier Koeffizienten hingegen ist deutlich kleiner und diese werden durch die FMM relativ genau berechnet.

In drei Dimensionen wird ein so genanntes Schachbrett-Beugungsgitter näher betrachtet. Die Erweiterungen der FMM in 3D werden miteinander verglichen und festgestellt, dass die Normalen-Vektor-Methode für dieses Beispiel deutlich bessere Ergebnisse liefert als die übliche FMM und die Jones Polarisationsbasis. Letztere hingegen liefert genaue Ergebnisse für das zweite Beispiel eines Kontaktlochs in einem absorbierenden Material. Die so genannte Methode der Subpixel Glättung hingegen zeigt zwar schnelle selbst-konsistente Konvergenz, deren Ergebnisse weichen aber stark von denen der anderen FMM-Varianten und den FEM Resultaten ab. Die Konvergenz der üblichen FMM ist auf Grund der unvollständigen Anwendung der inversen Regel deutlich langsamer. Deshalb werden für das letzte Beispiel einer photonischen Kristall-Schicht in 3D nur die Normalen-Vektor und die Jones Methode mit der FEM verglichen. Die Banddiagramme der ersteren zeigen unterschiedliches Verhalten zu FEM-Ergebnissen, insbesondere für wenige Fourier Basisfunktionen. Die Resultate der Jones Methode der FMM hingegen sind vergleichbar mit denen der FEM und die Fehler der Energieerhaltung sind sogar geringer verteilt.

Zusammenfassend zeigt die vorliegende Arbeit, dass die FMM für die Beugungseffizienzen dielektrischer Strukturen und die energetische Betrachtung bei photonischen Kristallen gute Ergebnisse liefert. Jedoch sind die Genauigkeiten der Nahfeld-Eigenschaften und die Phasenkorrelationen der Fernfeld Fourier Koeffizienten begrenzt. Deshalb kann die FMM dazu dienen vorläufige Untersuchungen und Simulationen durchzuführen. Sie sollte jedoch durch die FEM, deren numerische Fehler besser kontrolliert werden können, ergänzt werden.

Chapter I

Introduction

Nano-optical scattering problems play an important role in our modern, technologically driven society. Computers, smartphones and all kinds of electronic devices are manufactured by the semiconductor industry which relies on production using photomasks as well as optical process control. The digital world, e.g. the world wide web, is based on optical interconnects and so-called quantum computers based on optics are supposed to be next generation computers. Moreover, global economic progress demands new and sustainable energy resources and one option is to make use of the power stored in optical radiation from the sun. Additionally, understanding fundamental physics such as the optical properties of asymmetric, or chiral, structures could promote future innovations in engineering.

In order to understand and manipulate these kinds of processes, physics provides a well established model: the so-called Maxwell's equations. Stated by James Clerk Maxwell in 1862, this description of the interaction of light and matter still provides a profound basis for the analysis of electromagnetic phenomena. However, real world problems cannot be calculated using simple mathematics. Rather, computer simulations are needed to obtain solutions of the physical model.

Finding suitable methods to solve these problems opens up a wide variety of possibilities. On the one hand, there are methods which require long computing times. On the other hand, some algorithms depend on high memory usage. That is why the field of numerics deals with the question which method is optimally suited for specific problems.

The aim of this work is to investigate the applicability of the so-called Fourier Modal Method (FMM) to nano-optical scattering problems in general. Since simple analytical solutions are non-existent for most recent physical problems, we use the Finite Element Method (FEM) to double-check performance of the FMM. Mathematics provide reliable procedures to control the errors of numerics using the FEM. Yet up to now it has not been possible to rigorously classify the quality of the Fourier Modal Method's results. It is not fully understood whether the process of investing more and more computing resources yields more accurate results. So, we have to ask ourselves: does the numerical method invariably converge?

In spite of this uncertainty when using the FMM, it is a well established method dating back to the 1980s. This numerical method has recently been used to optimize performance of solar cells [19] as well as to improve the optical properties of so-called single-photon sources [41] which are essential for quantum cryptography. The latter is a promising candidate to increase digital security and revolutionise cryptography techniques. Furthermore, with the help of the Fourier Modal Method an important issue in optics has been partly resolved: angular filtering of light was made possible by using a mirror which becomes transparent at a certain viewing angle [77]. In addition, an improved numerical technique to design so-called Photonic Crystal waveguides based on the FMM was developed recently [15]. Photonic Crystals are used in the fields of optical bio-sensing and for the construction of novel semiconductor devices. Moreover, approaches to link the FMM and the FEM try to combine advantages of both methods to obtain fast and accurate results [81]. These ideas are closely linked to the well-known concept of Domain Decomposition within the FEM [88]. Here, one possibility to couple domains is to use the scattering matrix formalism as it is done in the FMM.

In the scope of this convergence study, we state Maxwell's equations, particularly for periodic geometries. We describe two physical phenomena of nano-optics, namely chirality and opto-electrical coupling, and define the errors of our simulations. Afterwards, the two investigated methods are analysed with respect to their general properties and a way to unify modelling physics when using both algorithms is presented. With the help of various numerical experiments, we explore convergence characteristics of the FMM and draw conclusions about the ability of this approach to provide accurate results and, consequently, its potential for research on technological innovations.

Chapter II

Background

In this chapter we state the background for this project. First we recapitulate the well established model of electrodynamics for simulations of nano-optical scattering problems and especially its frequency domain formulation. Since the FMM originates from early works on diffraction gratings a short introduction to this field is given afterwards. Two examples for recent interest in near-field simulations (optical chirality and opto-electrical coupling) are described in Section II.3. Finally, the errors investigated throughout this convergence study are motivated and defined.

II.1 Electrodynamics

II.1.1 Maxwell's Equations

Maxwell's equations are the basis for classical electrodynamics and are used as macroscopic as well as microscopic model for nano-optical scattering problems. In differential form they read [31]

$$\boldsymbol{\nabla} \times \boldsymbol{E} = -\partial_t \boldsymbol{B} \tag{II.1}$$
$$\boldsymbol{\nabla} \times \boldsymbol{H} = \partial_t \boldsymbol{D} + \boldsymbol{J} \tag{II.2}$$
$$\boldsymbol{\nabla} \cdot \boldsymbol{B} = 0 \tag{II.3}$$
$$\boldsymbol{\nabla} \cdot \boldsymbol{D} = \rho. \tag{II.4}$$

$\boldsymbol{E}$ and $\boldsymbol{H}$ are the electric and magnetic field strenghts, respectively. $\boldsymbol{D}$ is the so-called electric displacement field and $\boldsymbol{B}$ the magnetic flux density. The macroscopic charges ρ together with the macroscopic current density $\boldsymbol{J}$ fulfil the continuity equation $\partial_t \rho + \boldsymbol{\nabla} \cdot \boldsymbol{J} = 0$ which is of a form similar to a greater number of conserved quantities (see Sec. II.3.1). Equations (II.1) and (II.3) are the so-called homogeneous Maxwell's equations, whereas equations (II.2) and (II.4) are the inhomogeneous Maxwell's equations.

For the four vectorial quantities $\boldsymbol{E}$, $\boldsymbol{D}$, $\boldsymbol{H}$ and $\boldsymbol{B}$ there are twelve unknowns. Since there are only eight Maxwell's equations we need the following constitutive equations to relate the electric field and the electric displacement field and the magnetic field and the magnetic flux density

$$\boldsymbol{D} = \varepsilon \boldsymbol{E} \tag{II.5}$$
$$\boldsymbol{B} = \mu \boldsymbol{H} \tag{II.6}$$
$$\boldsymbol{J} = \sigma \boldsymbol{E}. \tag{II.7}$$

Here $\varepsilon = \varepsilon_0 \varepsilon_r$ is the permittivity consisting of the vacuum permittivity ε_0 and the relative permittivity ε_r. Similar definitions hold for the permeability μ. σ is the conductivity. These parameters are dependent on the materials and are usually of tensorical form but reduce to scalars for isotropic materials.

II.1.1.1 Time-Harmonic Formulation

In numerical computations one option is to solve Maxwell's equations in frequency domain. This means we make the ansatz

$$\boldsymbol{X} = \mathrm{Re}\left[\boldsymbol{\mathcal{X}} \exp\left(-i\omega t\right)\right] \tag{II.8}$$

for the physically observable real quantities $\boldsymbol{X}$, i.e. $\boldsymbol{E}$, $\boldsymbol{D}$, $\boldsymbol{H}$, $\boldsymbol{B}$ and $\boldsymbol{J}$. The complex quantities $\boldsymbol{\mathcal{X}}$ include phase information of the real valued quantities and the steady state frequency is ω. Maxwell's equations reduce then to

$$\boldsymbol{\nabla} \times \boldsymbol{\mathcal{E}} = i\omega\boldsymbol{\mathcal{B}} \tag{II.9}$$

$$\boldsymbol{\nabla} \times \boldsymbol{\mathcal{H}} = -i\omega\boldsymbol{\mathcal{D}} + \boldsymbol{\mathcal{J}} \tag{II.10}$$

$$\boldsymbol{\nabla} \cdot \boldsymbol{\mathcal{B}} = 0 \tag{II.11}$$

$$\boldsymbol{\nabla} \cdot \boldsymbol{\mathcal{D}} = \rho. \tag{II.12}$$

In the following only charge-free systems ($\rho = 0$) will be analysed. We use the constitutive equations (II.5)-(II.7) and redefine the permittivity tensor as a complex quantity $\varepsilon \to \varepsilon + i\sigma/\omega$. Taking the rotation of Eq. (II.9) yields

$$\boldsymbol{\nabla} \times \mu^{-1}\boldsymbol{\nabla} \times \boldsymbol{\mathcal{E}} - \omega^2\varepsilon\boldsymbol{\mathcal{E}} = 0. \tag{II.13}$$

II.1.2 Numerics on Maxwell's Equations

Technological progress in optics and electronics has lead to more and more complicated devices which cannot be handled with analytical solutions. On the other hand, increasing power and performance in numerics and computer science open up the possibility to solve Maxwell's equations numerically and study the properties of these structures, optimize geometrical or material parameters and gain insight into physical processes in the far and near electromagnetic fields. At the beginning of this evolution experimental data was confirmed with far-field approximations yet research focuses on microsopic effects as well.

There are various numerical techniques for different purposes. Geometrical optics can be studied using the so-called Beam Propagation Method which is capable of handling large devices. For periodic systems variations of the FMM are well established. FEM is a mathematically well studied method which can be used and optimized for a wide range of problems in nano-optics. Contrary to the ansatz of FMM and time-harmonic FEM (see Sec. II.1.1.1), Maxwell's equations can also be solved in time domain. The simplest and most common technique is the so-called Finite-Difference Time-Domain (FDTD) method. Here, the key is a discretized approximation of the differential operators in Maxwell's equations and its advantage is to compute many frequencies simultaneously. However, it lacks possibility of optimization and adaptivity. That is why advanced time domain methods such as the Discontinuous Galerkin Method [27], which is closely related to the FEM, are studied in more detail. Additionally, new formulations including the so-called Discontinuous Petrov-Galerkin Method have recently been proposed [12].

Although it is a challenge in itself to choose the right method for a specific problem, the aim of this project is to investigate the general applicability of the FMM for problems in the field of nano-optics. We attempt to come to general conclusions from the analysis of several examples of nano-optical devices. In order to double-check numerical results and to be able to argue on a well established convergence theory, simulations are compared to the FEM. However, it should be noted that custom formulations and implementations of the various methods mentioned above could possibly be much more suitable for very advanced and specialized cases.

II.2 Diffraction Theory

Since the origin of the FMM, which is in the focus of this work, lies in grating theory we shortly motivate periodic structures in the field of optics and especially diffraction gratings. We illustrate the basic idea of generalizing the Rayleigh expansion to non-homogeneous media which leads from Maxwell's equations to the formulation of the FMM (see Sec. III.1.4).

II.2.1 Grating Theory

Diffraction gratings are widely used for redirecting light in its spectral content. The invention of these gratings [69] is located in the use of periodicity on the scale of the wavelength of light. Devices of this theory are found e.g. in the field of spectroscopy by which circular dichroism is measured and can be manipulated by chiral structures (see Sec. II.3.1). Furthermore, gratings appear in astronomy, lasers and optical communication for instance as fiber grating couplers [84]. Modern variations of 1D periodic gratings are 2D and 3D periodic Photonic Crystals (PhCs) which show a band gap for light similar to electronic band gaps known from solid state physics.

For the following we assume periodicity in the x-direction with a pitch Λ. The grating number is defined as $K = 2\pi/\Lambda$. For an illumination with incident angle θ_i, Snell's law states the preservation of the wave vector at a material interface [58]: $k_x = k\sin(\theta_i)$. In grating theory the m-th diffraction order is defined with wave vector $k_{x,m} = k\sin(\theta_i) + mK$. These relations result in the famous Fraunhofer or grating equation for the diffraction angles

$$\sin(\theta_m) = \sin(\theta_i) + m\frac{\lambda}{\Lambda}. \tag{II.14}$$

The Fraunhofer equation accompanied by Kirchhoff's diffraction theory is sufficient for scalar optics. Here, only the direction of the diffraction order is of interest. However, if the period Λ is on the scale of the wavelength λ, grating theory needs to be extended to account for phenomena like Wood's anomaly [85]. This total absorption of light by a grating is explained with the help of the excitation of surface waves which are nowadays of high interest in the field of Surface Plasmon Polaritons [51]. It occurs due to an interplay of the so-called propagating diffraction orders with real k_x and the evanescent diffraction orders with imaginary k_x.

In grating theory a number of diffraction methods evolved in the past decades to simulate the results found in experiments. Mainly integral methods, modal methods and differential methods are used [58]. Each grating theory has its specific limits and altogether they are complementary for different gratings. Yet pros and cons of these methods are often qualitative and based on experiences. It is the aim of this work to quantify the validity of the FMM. Nevertheless, all these numerical approaches to solve Maxwell's equations for periodic structures make use of the particular periodic feature which is illustrated in the next section.

II.2.2 Periodic Electrodynamics in 2D

For homogeneous regions Rayleigh proposed to write the electromagnetic fields as a series of propagating and evanescent waves [68]

$$F(x,z) = \exp(ik\sin\theta_i x)\sum_m F_m(z)\exp(iKx), \tag{II.15}$$

with the incident angle θ_i, the grating number K and the unknown Fourier coefficients $F_m(z)$. For rectangular grooves in gratings, i.e. lamellar gratings, the coeffcients F_m are constant. This is the basis of the FMM. Alternatively, coordinate transformations can be used to simplify the boundary conditions at arbitrary surfaces. One of the most famous methods is the so-called C-method [14].

We use a linear operator $\mathcal{R}$ for a grating which relates the incoming fields F_{in} to the outgoing fields F: $F = \mathcal{R}F_{\text{in}}$. In doing so, we derive a property of the electromagnetic fields from the periodicity of the geometry: Since the grating is invariant for translations of the period Λ we obtain the so-called pseudoperiodicity of electromagnetic fields

$$F(x+\Lambda,z) = \exp(i\alpha_0\Lambda)\,F(x,z), \tag{II.16}$$

with $\alpha_0 = k\sin\theta$ and the diffraction angle θ. The function $V(x,z) = \exp(-i\alpha_0 x)\,F(x,z)$ is strictly periodic in x, so we can use a Fourier Transform (FT) to represent it: $V(x,z) = \sum_n u_n(z)\exp(inKx)$. This leads to the so-called pseudo-Fourier series of the electromagnetic fields

$$F(x,z) = \sum_n F_n(z)\exp(i\alpha_n x) \tag{II.17}$$

with $\alpha_n = \alpha_0 + nK$.

The ideas above pave the way to solve the time-harmonic Maxwell equations (II.9)-(II.12) in z-

invariant layers to obtain so-called eigenmodes. Structures are decomposed into layers (see Sec. III.3). It follows from the z-invariance for e.g. the electric field $\boldsymbol{E} = \boldsymbol{e}(\boldsymbol{r}_\perp)\exp(i\beta z)$ with the propagation constant β and the in xy-plane space vector $\boldsymbol{r}_\perp$. For simplicity we assume nonmagnetic $\mu = 1$ and charge-free $\rho = 0$ media. For these it follows from Gauss's law: $\boldsymbol{\nabla}\cdot\boldsymbol{D} = 0 \Rightarrow \boldsymbol{\nabla}\cdot\boldsymbol{E} = -\boldsymbol{E}\cdot(\nabla \ln\varepsilon)$. Using the z-invariance of ε, the eigenmodes fulfil the following equation [39]

$$\nabla_\perp^2 \boldsymbol{e}_\perp(\boldsymbol{r}_\perp) + \nabla_\perp\left(\boldsymbol{e}_\perp(\boldsymbol{r}_\perp)\cdot\nabla\ln\varepsilon(\boldsymbol{r}_\perp)\right) + \varepsilon(\boldsymbol{r}_\perp)k_0^2\boldsymbol{e}_\perp(\boldsymbol{r}_\perp) = \beta^2\boldsymbol{e}_\perp(\boldsymbol{r}_\perp) \tag{II.18}$$

with the vacuum wave number $k_0 = \omega/c_0$.

In 2D, i.e. for y-invariant geometries, this form of Maxwell's equations decouples into two sets of ordinary differential equations. These are the so-called transverse electric (TE) and transverse magnetic (TM) polarisations:

$$\partial_x^2 e_y + \varepsilon(x)k_0^2 e_y = \beta^2 e_y \text{ (TE)} \tag{II.19}$$

$$\varepsilon(x)\partial_x\frac{1}{\varepsilon(x)}\partial_x h_y + \varepsilon(x)k_0^2 h_y = \beta^2 h_y \text{ (TM)}. \tag{II.20}$$

Often TE polarization is also called s-polarization and TM is called p-polarization.

The electric field in the layers is a series of forward (a_m) and backward (b_m) propagating modes

$$\boldsymbol{E}(\boldsymbol{r}_\perp) = \sum_m a_m \boldsymbol{e}_{\perp,m}(\boldsymbol{r}_\perp)\exp(i\beta_m z) + \sum_m b_m \boldsymbol{e}_{\perp,m}(\boldsymbol{r}_\perp)\exp(-i\beta_m z)\,. \tag{II.21}$$

The eigenmodes need to form a complete basis set which is only proven for dielectrics, i.e. $\varepsilon \in \mathbb{R}$ [71]. For metals and absorbing materials the completeness is usually simply assumed (see Sec. IV.2.3). To handle the layering and the z-invariance of the eigenmodes, transfer matrix algorithms ensure the boundary conditions at the layer interfaces. While the so-called T-matrix formalism relates forward to backward propagating fields, the S-matrix algorithm connects outgoing and incoming fields and is used in the FMM because it is numerically stable [42]. Both transfer matrix formalisms make use of the Fresnel coefficients following from Snell's law:

$$r_s = \frac{E_{r,s}}{E_{i,s}} = \frac{n_i\cos(\theta_i) - n_t\cos(\theta_t)}{n_i\cos(\theta_i) + n_t\cos(\theta_t)} \tag{II.22}$$

$$t_s = \frac{E_{t,s}}{E_{i,s}} = \frac{2n_i\cos(\theta_i)}{n_i\cos(\theta_i) + n_t\cos(\theta_t)} \tag{II.23}$$

$$r_p = \frac{E_{r,p}}{E_{i,p}} = \frac{n_i\cos(\theta_t) - n_t\cos(\theta_i)}{n_t\cos(\theta_i) + n_i\cos(\theta_t)} \tag{II.24}$$

$$t_p = \frac{E_{t,p}}{E_{i,p}} = \frac{2n_i\cos(\theta_i)}{n_i\cos(\theta_t) + n_t\cos(\theta_i)}, \tag{II.25}$$

where r_s, t_s and r_p, t_p are the reflection and transmission coefficients of s- and p-polarization, respectively. n_i, θ_i and n_t, θ_t are the refractive index and the diffraction angle of incidence and transmittance, respectively. The incoming electric field is E_i, the reflected one E_r and the amplitude of the transmitted field is E_t.

In the FMM the modes $\boldsymbol{e}_{\perp,m}$ are expanded on a Fourier series motivated by the pseudoperiodicity of the fields (II.17). An alternative is the so-called semianalytical approach. It is based on analytically known eigenmodes for e.g. rotational symmetric structures [39]. Here, the propagation constants of guided and semi-radiation modes are determined with a method-specific search routine in the complex plane.

II.2.3 Diffraction Efficiency

The major interest of diffraction theory are the grating or diffraction efficiencies. These are defined as the reflected or transmitted fluxes Φ_m of each diffraction order normalized with the incoming flux Φ_i:

$$e_m^{(t,r)} := \frac{\Phi_m}{\Phi_i}. \tag{II.26}$$

For propagating plane waves the time harmonic Poynting vector is [31]

$$\boldsymbol{S} := \frac{1}{2}\boldsymbol{E} \times \boldsymbol{H}^*. \tag{II.27}$$

For these plane waves it holds $|\boldsymbol{H}| = |\boldsymbol{E}|/Z$ with the wave impedance Z. Since the flux through a surface $\mathcal{F}$ is $\Phi = \int_{\mathcal{F}} \boldsymbol{S} d\boldsymbol{f}$ it follows for the m-th diffraction order

$$\Phi_m = |\boldsymbol{S}| \cos(\theta) \tag{II.28}$$

with the diffraction angle θ. That is why we obtain for the diffraction efficiencies in terms of the Fourier coefficients $\boldsymbol{E}_m^{(t,r)}$

$$e_m^{(t,r)} = \frac{\cos(\theta_m)|\boldsymbol{E}_m^{(t,r)}|^2 n_{(t,i)}}{\cos(\theta_i)|\boldsymbol{E}_i|^2 n_i} \tag{II.29}$$

with the incoming field $\boldsymbol{E}_i$, incident angle θ_i and the refractive indices in the region of transmittance and incidence $n_{(t,i)}$, respectively.

From energy conservation in lossless media follows the so-called energy-balance criterion: $\sum e^{(r)} + \sum e^{(t)} = 1$. In grating theory this is the main condition which should be fulfilled by a numerical method. That is why in this field the following error needs to be small when talking of convergent results [58]

$$\frac{1}{N}\sum_n \frac{e_n^{(t,r)} - \tilde{e}_n^{(r,t)}}{e_n^{(t,r)}}, \tag{II.30}$$

where N is the number of propagating modes, $e_n^{(t,r)}$ the correct diffraction efficiency and $\tilde{e}_n^{(t,r)}$ the numerical result. This error is similar to the one defined in Definition II.4.9 and will be subject of this work. However, for modern grating couplers and in the field of metrology the correct far-field pattern needs to be resolved on a nanometre scale. That is why we also analyse the phase correlations of the Fourier coefficients (Def. II.4.7).

II.3 Near-Field Effects

II.3.1 Optical Chirality

Recent developments of nanophotonics are driven by numerical simulations. Experimental as well as theoretical physicists hope to obtain more insight into underlying physical processes by modelling and simulating them. That is why near-field computation is a key feature of modern simulation tools. In Chapter IV we investigate characteristics of near-field results for FEM and FMM. In the following we outline one example of great interest in this visualization of electromagnetic fields on a nanometre scale. Optical chirality follows from the dual symmetry of electric and magnetic fields described and generalized in Appendix B.

II.3.1.1 Motivation

The physics of molecules is partly understood via circular dichroism (CD) which describes the differential absorption of left- and right-circularly polarized light (CPL). Chiral molecules cannot be superimposed with their mirror image and are highly sensitive to CD. In 2010 Tang and Cohen [80] proposed using a quantity introduced in 1964 by Lipkin [47] for measuring the chirality of an electromagnetic field:

$$C := \frac{\varepsilon_0}{2}\boldsymbol{E} \cdot \boldsymbol{\nabla} \times \boldsymbol{E} + \frac{1}{2\mu_0}\boldsymbol{B} \cdot \boldsymbol{\nabla} \times \boldsymbol{B}. \tag{II.31}$$

Lipkin called this time-even pseudoscalar “zilch” and supposed that it had no physical meaning at all. It is conserved in vacuum as Lipkin showed and commonly used in recent publications [74].

Classifying properties of the electromagnetic field according to their symmetries reveals a missing quantity analogous to parity in particle physics which is scalar, antisymmetric under mirror reflection and symmetric under time reversal (Fig. II.1).

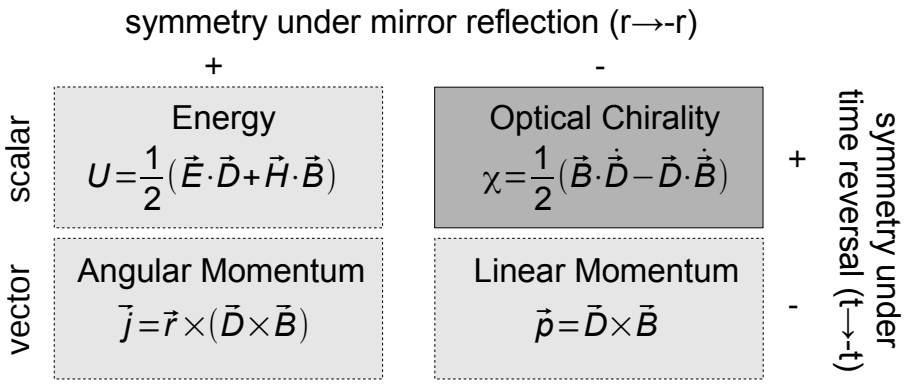


Figure II.1: *Symmetry behaviour of electromagnetic conserved quantities. Standard quantities include energy, angular momentum and linear momentum. Analysing transformation behaviour of these conservation laws under mirror reflection (columns) and time reversal (rows) reveals a missing scalar quantity which is odd (-) under mirror reflection and even (+) under time reversal. A candidate for such a quantity is the so-called optical chirality introduced by Tang and Cohen [80].*

II.3.1.2 Chirality in Medium

Tang and Cohen use a standard formula for the rate of excitation of molecules [26] to conclude their representation of chirality and connect the resulting quantity in SI units to Lipkin's "zilch" in Gaussian units. The resulting quantity has SI units of force density which seems to be counter-intuitive since it is rather a scalar than a vectorial quantity as force. Furthermore, both Tang and Cohen and Lipkin defined chirality only in electromagnetic vacuum. Ragusa and Baylin [67] defined zilch in a medium and showed in the context of electromagnetic field theory that this zilch is only conserved for media satisfying $\varepsilon\mu = 1$.

Philbin [62] analysed Lipkin's conservation law with Noether's theorem and identified a transformation of the classical electromagnetic vector potential resulting in zilch's conservation. Analogous to his definition of optical zilch and the corresponding flux in non-dispersive media we define chirality density χ and chirality flux density $\boldsymbol{\Sigma}$ respectively:

$$\chi := \frac{1}{2}\left(\boldsymbol{B}\cdot\dot{\boldsymbol{D}} - \boldsymbol{D}\cdot\dot{\boldsymbol{B}}\right) \tag{II.32}$$

$$\boldsymbol{\Sigma} := \frac{1}{2}\left(\boldsymbol{E}\times\dot{\boldsymbol{D}} + \boldsymbol{H}\times\dot{\boldsymbol{B}}\right). \tag{II.33}$$

The continuity equation for chirality follows from Maxwell's equation (II.1)-(II.4) in current and charge-free space ($\boldsymbol{J} = 0$, $\rho = 0$):

$$\dot{\chi} + \boldsymbol{\nabla}\cdot\boldsymbol{\Sigma} = 0. \tag{II.34}$$

In the time-harmonic context we define complex chirality density $\mathfrak{X}$ and complex chirality flux density $\mathfrak{S}$:

$$\mathfrak{X} := -\frac{1}{2}i\omega\boldsymbol{\mathcal{B}}^*\cdot\boldsymbol{\mathcal{D}} \tag{II.35}$$

$$\mathfrak{S} := -\frac{1}{4}i\omega(\boldsymbol{\mathcal{H}}^*\times\boldsymbol{\mathcal{B}} + \boldsymbol{\mathcal{E}}^*\times\boldsymbol{\mathcal{D}}). \tag{II.36}$$

Real parts of the quantities defined above correspond to time-averaged values of (II.32) and (II.33) [cf. ansatz (II.8)]. These definitions yield results similar to currently used formulae for time-harmonic chirality in vacuum [73].

II.3.1.3 Chiral Energy Density

Despite being investigated recently, chirality is far from being fully understood. There is no comparable physical quantity with dimensionality of force density. Chirality is sometimes analysed in correlation with energy: Bliokh and Nori outline symmetries between continuity of chirality and Poynting's theorem [10]. Tang and Cohen generalize the dissymmetry factor g of a monochromatic CPL to

electromagnetic fields interacting with chiral molecules. They are using a factor of $2C/(\omega u_e)$ with the electric energy density u_e [80]. Philbin analyses zilch per unit energy and concludes that "zilch flows through the medium at the group velocity $c/n_g(\omega)$, just like the optical energy" [62].

Following these ideas we define chiral electric $\boldsymbol{E}_\chi$ and magnetic fields $\boldsymbol{H}_\chi$ with the help of the wave impedance $Z = \sqrt{\mu/\varepsilon}$:

$$\boldsymbol{E}_\chi := \sqrt{\frac{\mu}{\varepsilon}}\boldsymbol{H} \tag{II.37}$$

$$\boldsymbol{H}_\chi := \sqrt{\frac{\varepsilon}{\mu}}\boldsymbol{E}. \tag{II.38}$$

For the chiral energy density

$$\widetilde{\chi} := \frac{1}{2\omega}\left[\boldsymbol{E}_\chi \cdot (\boldsymbol{\nabla}\times\boldsymbol{H}) - \boldsymbol{H}_\chi \cdot (\boldsymbol{\nabla}\times\boldsymbol{E})\right] \tag{II.39}$$

and the corresponding chiral energy flux density

$$\widetilde{\boldsymbol{\Sigma}} := -\frac{1}{2\omega}\left[\boldsymbol{H}_\chi \times \varepsilon^{-1}(\boldsymbol{\nabla}\times\boldsymbol{H}) + \boldsymbol{E}_\chi \times \mu^{-1}(\boldsymbol{\nabla}\times\boldsymbol{E})\right] \tag{II.40}$$

the continuity equation (II.34) is still satisfied. Their time-harmonic averaged values are

$$\widetilde{\mathfrak{X}} := -\frac{1}{4}i\left(\boldsymbol{E}_\chi^* \cdot \boldsymbol{D} + \boldsymbol{H}_\chi^* \cdot \boldsymbol{B}\right) \tag{II.41}$$

$$\widetilde{\boldsymbol{\Sigma}} := -\frac{1}{4}i\left(\boldsymbol{E}\times\boldsymbol{H}_\chi^* + \boldsymbol{E}_\chi^* \times \boldsymbol{H}\right). \tag{II.42}$$

Although the proposed redefinition of chirality as energy is only a frequency scaling at first sight, it could help in understanding chirality in general:

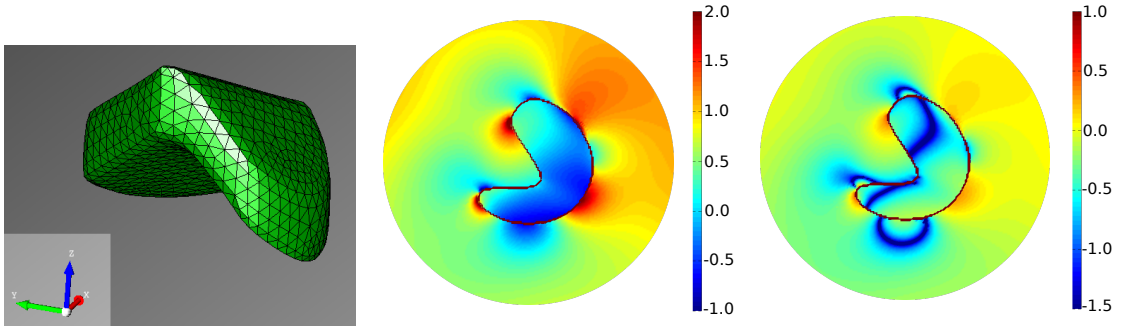
1. First of all, the physical interpretation of splitting energy in chiral and non-chiral parts is comparable to splitting energy in s- and p-polarized energy parts in diffraction theory.
2. Secondly, the definition of (II.39) is consistent with the exterior calculus used for time-harmonic analysis of Maxwell's equations [89]. Chiral energy density is a proper 3-form and chiral energy flux is a 2-form as expected from the physical interpretation of mathematics.
3. Thirdly, the time-averaging of the proposed quantities (II.41) is in close connection to the form of time-averaged energy density and the Poynting vector.
4. Additionally, the numerical problem of an order of magnitude mismatch between chirality and energy (supplementary material of [73]) is resolved.

Using this picture of chiral and non-chiral energy one could develop a formalism for media interface behaviour of chirality analogous to the well-known transfer or S-matrix formalisms of s- and p-polarization [39]. As a result, the recently proposed Chiral Jones Matrix [17] may be generalized.

II.3.1.4 Chirality of Nanoparticles

Currently, a variety of chiral and achiral structures are investigated with respect to their electromagnetic chiral properties and their use for tailoring enhanced CD [74]. One can either excite a chiral device with an achiral source, vice versa, or, as it is more commonly done, excite chiral geometries with CPL. The latter are optical chirality eigenstates with constant chirality in vacuum [62]. Controlling locally enhanced chirality paves the way to understanding and designing interaction between chiral molecules and electromagnetic fields. The highest recorded molar CD is obtained with chiral nanoparticles [Fig. II.2(a)] in the far-field [53]. Analysing near-field behaviour possibly helps further design of improved structures and enhances their coupling to optical dipole sources.

Exciting these metallic nanoparticles with CPL yields a complex chirality density field. For the following the generalization of the commonly accepted chirality density (II.35) is used. Chirality of the incoming CPL is constant while in the region of the particle it is enhanced and its sign is changed as well [Fig. II.2(b)]. The interfaces of sign changes are clearer on a logarithmic scale [Fig. II.2(c)]. Here,



(a) Geometry (tetrahedral mesh) (b) Chirality enhancement (linear) (c) Chirality enhancement (ln)

Figure II.2: *Chiral metallic nanoparticle (a). For details about this structure refer to [53] and L. Poulikakos*[1]*. Chirality enhancement over homogeneous media is plotted on linear scale (b) and natural logarithmic scaling (c) of absolute value in a z [blue axis in (a)] cross section of the particle. Particle's shape is depicted by red lines. Illumination is* $-k_y(R)$ *(see Fig. II.3 for details). Compared to former work [74] not only chirality enhancement outside of but also inside the particle is shown with the help of the formalism developed in Section II.3.1.2. A widespread sign change in the particle can be observed (b) and the isolines of sign change are more visible in logarithmic scaling (c). We suspect that these sign changes could be a key in understanding chirality enhancement and could be used for designing particles which show higher electromagnetic chirality.*

one can see a chirality density structure inside and outside of the particle and the close connection of these interior and exterior patterns. Due to a missing general formula, previous publications refer only to the vacuum part of the computational domain (CoDo).

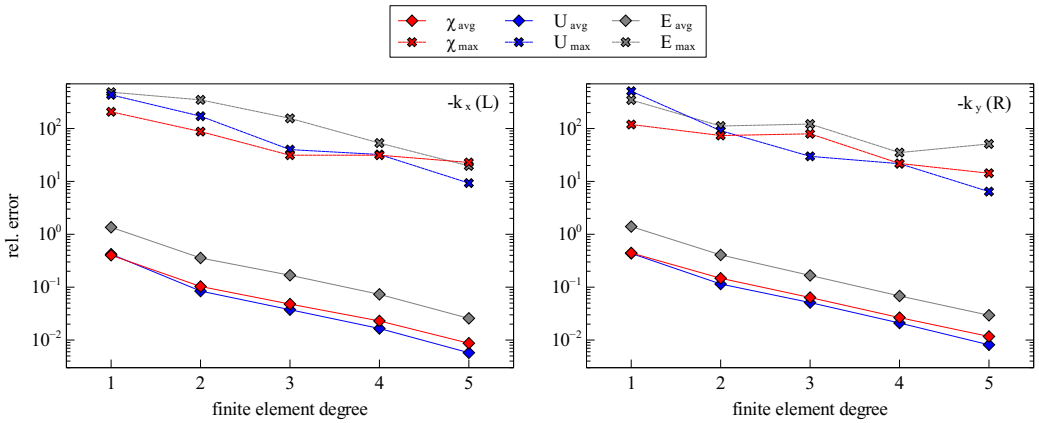
Further investigation could reveal the interplay of chirality at material interfaces as well as the missing interpretation of the complex part of (II.32). The latter may be related to absorbed chiral energy in comparison with the complex part of the time-averaged electric energy.

For this near-field analysis, a converged field pattern is the key, as mentioned at the beginning of this section. Although point evaluation is badly conditioned in FEM, Figure II.3(a) shows the convergence of an equally spaced $80 \times 80 \times 80$ Cartesian grid on a fixed mesh for increasing polynomial degree. Relative errors are plotted with respect to the highest polynomial degree ($p = 6$) and in pointwise absolute values. The average error is computed on the equally spaced grid and its maximum is displayed as well. In Figure II.3(b) the error of integral quantities is shown via a density integration of the whole CoDo. As expected, integral values converge much faster due to the weak formulation used within FEM (see Sec. III.2.1).

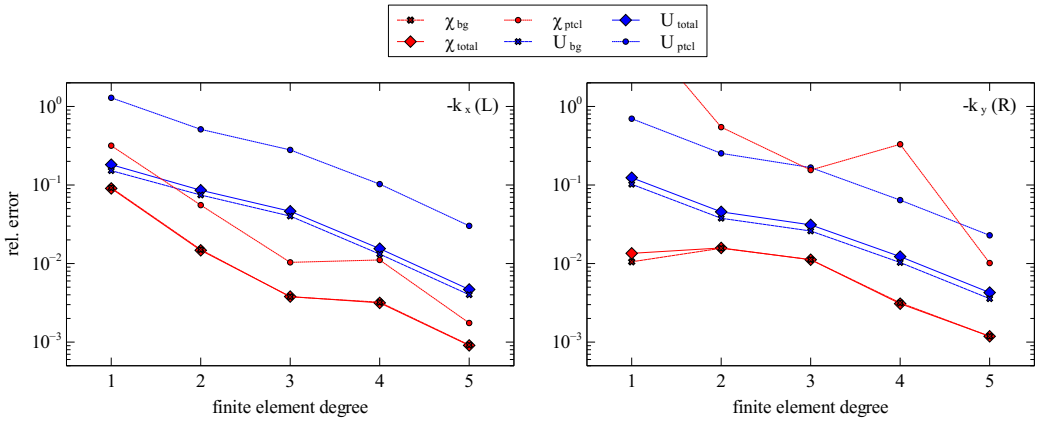
Nevertheless, this convergence analysis shows that convergence characteristics of chirality are not as smooth as those for the electric field energy. This can be understood because of the fact that for the electric field energy only the electric field needs to be computed. On the other hand, for chirality (II.35) magnetic fields are needed as well and are computed as numerical derivatives of the electric field. Convergence could be enhanced by solving the corresponding electric and magnetic fields separately and combining numerical results independently in the framework of *JCMsuite* [32].

Additionally, convergence of the near-field as well as the integrated quantities depends heavily on the direction and polarization of illumination. Adaptive strategies (see Sec. III.2.4) are promising candidates to obtain more stable results in 3D simulations.

[1]Lisa Poulikakos. PhD Candidate, ETH Zürich, Optical Materials Engineering Lab, Switzerland. E-Mail: plisa@ethz.ch.



(a) Pointwise convergence on Cartesian grid. Average and maximum values are displayed.



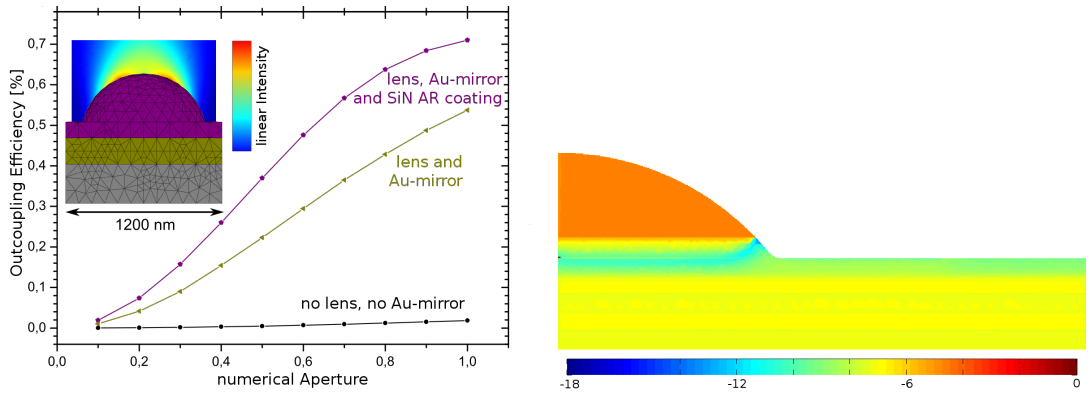
(b) Integral convergence via density integration. Quantities in the particle (ptcl), the sourrounding (bg) and in the total CoDo are displayed.

Figure II.3: *Convergence of FEM simulations of metallic nanoparticles with respect to finite element degree p. Errors of integrated chirality density and electric energy density are compared to the best simulation with $p = 6$. Two different directions of illumination are shown with respect to their k-vector (in the x- and the y-direction). Left- and right-handed CPL is denoted by L and R, respectively. Near-field chirality density shows larger errors than electric field energy density (a) because of the need for the magnetic field obtained by numerical derivation of the electric field. For the same reason integrated chirality shows nonsmooth convergence behaviour in contrast to electric field energy (b). Furthermore, this analysis shows that convergence of the figures of merit crucially depends on the direction of illumination and the polarization. That is why for anisotropic simulations averaging over a wide range of incident directions FEM parameters need to be chosen carefully when wanting to achieve the high accuracy needed for CD.*

II.3.2 Opto-electrical Coupling

A second example for interest in optical near-field behaviour originates from research on Single-Photon Sources (SPSs). The semianalytical approach of the FMM (see end of Sec. II.2.2) has been intensively used for tuning the optical properties of micropillars and photonic nanowires [41]. Even limiting effects of surface roughness for the Q-factor are studied with a cylindrical version of the Modal Method using the so-called staircase approximation [24]. The latter will be shortly analysed in Section IV.3.2. For these purposes scalar permittivity profiles are used. The FMM, however, is also capable to handle anisotropic materials [33]. Nevertheless, sophisticated implementation is needed for anisotropies in the z-direction and some 3D FMM formulations (see Sec. III.1.4) cannot generally handle all forms of tensorical permittivities [49].

Recent progress in the fabrication of deterministic SPS devices [25] motivates fully localized tensorical treatment of materials which currently is impossible for the FMM but naturally part of the FEM. As an example we mention a Quantum Dot of which both the spontaneous emission enhancement with respect to the Purcell Factor as well as the far-field directional emission in terms of outcoupling efficiency are optimized using FEM simulations [76]. Various setups are investigated with respect to the collected power into a certain numerical aperture [Fig. II.4(a)]. These include embedding the Quantum Dot in a simple planar substrate, placing a hemispherical lens on top and a gold mirror below the Quantum Dot and coating the latter with an antireflection coating. Optical near-field studies include testing the solution of vectorial Maxwell's equations against focal points of the lens computed with the help of geometrical optics. That is why it is worth investigating near-field convergence of numerical methods used in nanophotonics.



(a) Outcoupling efficiency

(b) Locally varying $\mathrm{Im}[\varepsilon(\boldsymbol{r})]$

Figure II.4: *Opto-electrical coupling for hemispherical lens on top of Quantum Dot. A variety of setups is under investigation in order to obtain high-performance SPS (a) (adapted from [76] on the authority of M. Seifried). These include embedding a Quantum Dot in a simple planar substrate (no lens, no Au-mirror), placing a gold mirror below and a hemispherical lens on top of the Quantum Dot (lens and Au-mirror) and coating the latter structure with an antireflection coating (lens, Au-mirror and SiN AR coating). Outcoupling efficiency with respect to the numerical aperture is obtained in FEM simulations. Near-field behaviour, such as focal points of the lens, are analysed. Additionally, the electrical properties of the SPS device can be simulated with a correction to the optical permittivity with the charge carrier densities (II.43) [5]. This mostly yields an additional local complex part to the permittivity field (b) (charge carrier densities simulated using the software package WIAS-TeSCA [82] by M. Kantner[2]). Tensorical permittivity can be directly included in FEM simulations but needs sophisticated improvement of the FMM [49].*

As noted before, optical effects of these kinds of devices were studied in detail with the FMM. To further tune the performance of electrically driven Quantum Dots, the approach of opto-electrical coupling could be beneficial: here, the influence of electrical effects on the optical model and vice versa are studied. A simplified iterative model is to couple local charge carrier densities to the optical refractive index. In this example the electron density $\mathfrak{n}$ and the hole density $\mathfrak{p}$ are computed with the software package WIAS-TeSCA [82]. Here the so-called drift diffusion model [5] is used to obtain the non-resonant background absorption α_{bg}. Then a correction to the purely optical refractive index

[2]Markus Kantner. PhD candidate, Weierstrass Institute for Applied Analysis and Stochastics Berlin, Laserdynamics, Germany. E-Mail: kantner@wias-berlin.de

$n(\omega)$ is computed yielding the local optical permittivity

$$\varepsilon(\omega, \boldsymbol{r}) = \left[n(\omega, \boldsymbol{r}) - ic_0 \frac{\alpha_{\mathrm{bg}}(\boldsymbol{r})}{2\omega} \right]^2, \tag{II.43}$$

where $\alpha_{\mathrm{bg}} = f_{\mathfrak{n}}\mathfrak{n}(r) + f_{\mathfrak{p}}\mathfrak{p}(r)$ with constants $f_{\mathfrak{n}} = f_{\mathfrak{p}} = 10^{-18}$ cm^2. This correction mostly leads to an increased absorption, i.e. imaginary part of ε. As an example of this locally varying absorption the upper part of a hemispherical lens on top of a Distributed Bragg Reflector (DBR) is displayed in Figure II.4(b). The lens is partly coated with a gold contact (to approximately half of the horizontal upper plane). For these results an external voltage of 1.5 V is applied to the gold contact leading to small absorption which drastically varies locally. The plot is cylindrically symmetric and on $\log_{10}$ scaling.

Although there are attempts to extend the FMM to spatial properties [49], it is much easier to cope with fully tensorical permittivity fields in the context of FEM. This demonstrates the limits of numerical methods in nanophotonics and recalls that certain methods are only suitable for special cases. These specialized methods such as the FMM the origins of which are in grating theory perform much better for limited research interests, but lack the generality of concepts such as the FEM. It should be noted that the original FMM only deals with plane wave illumination in scattering problems, but its extension to dipole sources needed for these kinds of simulations is currently used and studied [39]. Despite the fact that the Fourier representation of dipole sources is an interesting field of research, it is beyond the scope of the study at hand.

II.4 Error Notation

Since it is the aim of this project to analyse the convergence behaviour of the numerics on Maxwell's equations, it is crucial to decide how to measure errors. The convergence theory of FEM (see Sec. III.2.3) is formulated in the context of integral norms of the field distribution while the convergence theory of FMM (see Sec. III.1.2.3) deals with the pointwise approximation of the desired field itself. Corresponding near-field errors are defined in Section II.4.1.

Experimentalists are also interested in the far-field behaviour of their setups. To measure convergence for these, one needs the Fourier coefficients of the computed field. These different errors are defined in Section II.4.2. First of all, we define the absolute value of a complex vector. In the following, we use the term *absolute value* as synonym for the *norm* of a complex vector.

Definition II.4.1 (Absolute Value)**.** *Let $\boldsymbol{u} \in \mathbb{C}^n$. The absolute value of $\boldsymbol{u}$ is defined similar to the norm of $\boldsymbol{u}$:*

$$|\boldsymbol{u}| := \left(\sum |u_i|^2 \right)^{1/2}, \tag{II.44}$$

where for $u_i \in \mathbb{C}$

$$|u_i|^2 = \mathrm{Re}\,(u_i)^2 + \mathrm{Im}\,(u_i)^2\,. \tag{II.45}$$

II.4.1 Near-Field

II.4.1.1 Definitions

In most application cases one is interested in a specific figure of merit. But to draw general conclusions about the error behaviour of a numerical method, one should be interested in the approximation of the field itself. That is why we define the following norms which are of interest in the scope of this work.

Definition II.4.2 (p-Norm, L^p)**.** *Let $f(x)$ be an arbitrary function. The p-norm of f is*

$$\|f\|_p := \left(\int |f(x)|^p dx \right)^{1/p}. \tag{II.46}$$

The $p = \infty$ norm of f is defined as

$$\|f\|_\infty := \max_x |f(x)|. \tag{II.47}$$

The space L^p is the space of functions satisfying $\|f\|_p < \infty$. For numerical computations in the following chapters the region of integration is mostly restricted to the computational domain (CoDo).

In the following numerics these continuous norms become their discretized equivalences $\|f\|_p = (\sum_i |f(x_i)|^p)^{1/p}$ and $\|f\|_\infty = \max_{x_i} |f(x_i)|$, respectively.

In addition, we deal with the approximation u_h of the analytic solution u of Maxwell's equations. To classify the convergence of $u_h \to u$ we define the following normalized errors.

Definition II.4.3 (Relative L^p Error)**.** *Let u be the desired solution and u_h its approximation. The relative L^p error is defined as*

$$\Delta_{L^p} := \frac{\|u - u_h\|_p}{\|u\|_p}. \tag{II.48}$$

The Δ_{L^p} error of the field itself is computed throughout this work in its discretized form on a further described grid of points. For different finite element polynomial degrees in FEM one could interpolate between the solutions u and u_h. Yet for the comparison of FMM and FEM unnecessary complex integrals would have to be computed. That is why we use the discretized integrals mentioned above. In order to identify local contributions to Δ_{L^p} we define the following local error.

Definition II.4.4 (Local relative L^p Error)**.** *Let u be the desired solution, u_h its approximation on the CoDo Ω and $x \in \Omega$. The local relative L^p error is defined as*

$$\Delta_{L^p}(x) := \frac{|u(x) - u_h(x)|^p}{\max\limits_{y\in\Omega} |u(y) - u_h(y)|^p}. \tag{II.49}$$

Often one is interested in the electromagnetic energy related to the following integral error. Additionally, the following quantity can be computed more easily than the relative L^p error, since one can use the basis function of the numerical method used, e.g. plane waves for the FMM and polynomial functions for FEM. That is why there is a continuous and a discretized version of this error.

Definition II.4.5 (Relative Integral Error)**.** *Let u be the desired solution and u_h its approximation. The relative I^p error is defined as*

$$\Delta_{I^p} := \left| \frac{\|u\|_p - \|u_h\|_p}{\|u\|_p} \right|. \tag{II.50}$$

Let the continuous version of this error be $\Delta_{I^p}^{(c)}$: here, one uses the basis functions of the numerical method itself rather than evaluating functions on a grid of points.

In general we expect

$$\Delta_{I^p} \leq \Delta_{L^p} \tag{II.51}$$

which is demonstrated with the help of a simple example in Section II.4.1.2.

As mentioned above, the electric energy density is of major physical interest. It is not in general proportional to the values provided by the Δ_{I^2} error. For example in Section IV.3.2 we will investigate the so-called staircase approximation of FMM which inherently changes the modelled geometry and by that the local dependency of the permittivity ε which causes the non-proportionality. These are the reasons for the following definition.

Definition II.4.6 (Relative Energy Error)**.** *Let U be the desired energy of the CoDo for the investigated problem and U_h its approximation. Its relative error is defined as*

$$\Delta_U := \left| \frac{U - U_h}{U} \right|. \tag{II.52}$$

Let the continuous version of this error be $\Delta_U^{(c)}$: here, one uses the basis functions of the numerical method itself rather than evaluating functions on a grid of points.

Note that $U \in \mathbb{C}$ because we are dealing with time-harmonic representation of the electromagnetic fields. So one could also analyse the error of the propagating energy $\Delta_{\mathrm{Re}(U)}$ and the absorbed energy $\Delta_{\mathrm{Im}(U)}$, respectively.

II.4.1.2 Remark

As stated in the previous section, one is mainly interested in the convergence of specific quantities. Conclusions are often drawn from the convergence of these quantities and the convergence of the field distribution itself, e.g. physical effects are explained with the help of the near-field pattern for a new physical phenomenon. To emphasize the difference between the error of the intensity $|E|^2$ of a field E and the field approximation itself, the following basic example shows the derivation between Δ_{L^p} and Δ_{I^p} for $p = 1, 2$.

Let the field E and its approximation E_h be

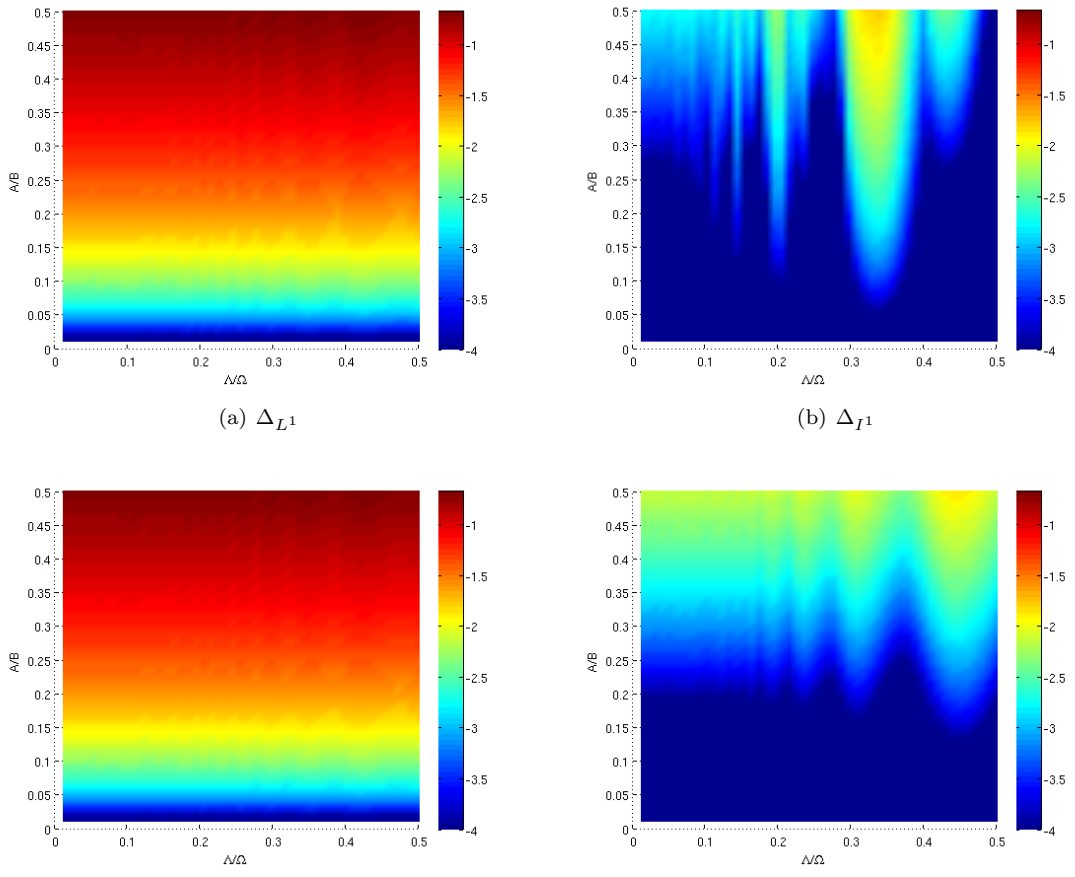
$$E(x) = B \sin\left(\frac{2\pi}{\Omega} x\right) \tag{II.53}$$

$$\delta(x) = A \sin\left(\frac{2\pi}{\Lambda} x\right) \tag{II.54}$$

$$E_h(x) = E(x) + \delta(x). \tag{II.55}$$

This is an easy model problem of the following since we compute periodic fields with a Fourier basis. For small and high frequency perturbations with small A and small Λ/Ω the integral error Δ_{I^p} estimates the error of the field approximation Δ_{L^p} some magnitudes smaller (Fig. II.5). Examples for a high and a low frequency perturbation are shown in Figure II.6. In Table II.1 the error values are displayed. Here, it can be seen that although the low frequency approximation shows the correct field pattern its Δ_{I^1} error is two magnitudes bigger on a logarithmic scale. Similar observations apply for $p = 2$.

The intention of this remark is to show the sensitivity of numerical convergence statements to the investigated error and the error estimation in use. This particularly applies for larger perturbations.



(a) Δ_{L^1} (b) Δ_{I^1}

(c) Δ_{L^2} (d) Δ_{I^2}

Figure II.5: *Errors of the model function (II.53) for small and high perturbations with varying periods. Natural logarithms of the errors (II.48) and (II.50) for $p = 1, 2$ are shown on a colorbar scale. Their different behaviour with respect to different perturbations can be seen, as well as the common underestimation of the field approximation with the integral error Δ_{I^p}.*

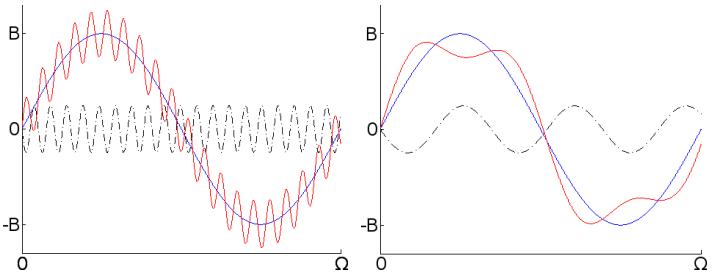


Figure II.6: *High (left) and low (right) frequent perturbations [black dashed line, (II.54)] of sinusoidal signal [blue solid line, (II.53)] and the total perturbed signal [red solid line, (II.55)].*

Table II.1: *Error estimations of the high and low frequent perturbations with errors (II.48) and (II.50).*

	high	low
Λ/Ω	0.051	0.347
Δ_{L^1}	0.244	0.250
$\ln(\Delta_{L^1})$	-1.410	-1.390
Δ_{I^1}	0.014	0.080
$\ln(\Delta_{I^1})$	-4.284	-2.555

II.4.2 Far-Field

For far-field comparison the Fourier Transform is used since it represents propagating plane waves. From the numerical solution of Maxwell's equations we obtain the approximation of the vector valued complex Fourier coefficients $\boldsymbol{f}_h^{(i)}$ of the analytical values $\boldsymbol{f}^{(i)}$ with the diffraction order $i \in \mathbb{Z}$. Similar to the formalism of the previous section we define the error of this approximation.

Definition II.4.7 (Relative Fourier Error)**.** *Let $\boldsymbol{f}^{(i)} \in \mathbb{C}^3$ be the desired Fourier coefficients and $\boldsymbol{f}_h^{(i)} \in \mathbb{C}^3$ their approximation. The relative Fourier error is defined as*

$$\Delta_F^{(i)} := \left| \frac{\boldsymbol{f}^{(i)} - \boldsymbol{f}_h^{(i)}}{\boldsymbol{f}^{(i)}} \right| \in \mathbb{R}_0^+, \tag{II.56}$$

where the division is meant by each coordinate separately: $\boldsymbol{a}/\boldsymbol{b} := (a_1/b_1, a_2/b_2, a_3/b_3)^T$. The total relative Fourier error is

$$\Delta_F := \frac{1}{N_f} \sum_i \Delta_F^{(i)}, \tag{II.57}$$

where N_f is the number of diffraction orders of the specific periodic problem.

Additionally, we define an error analogous to Δ_{L^∞} for the Fourier coefficients. We need this definition for symmetric devices where we expect zero values for certain components of the Fourier coefficient. Due to numerical errors these might not be exactly zero. If this is the case the value of the error defined above ($\Delta_F^{(i)}$) would be dominated by these numerically non-zero elements. To avoid this behaviour for e.g. the pin hole of Section IV.4.2 we define:

Definition II.4.8 (Relative maximal Fourier Error)**.** *Let $\boldsymbol{f}^{(i)} \in \mathbb{C}^3$ be the desired Fourier coefficients and $\boldsymbol{f}_h^{(i)} \in \mathbb{C}^3$ their approximation with the Cartesian components $\left(\boldsymbol{f}_h^{(i)}\right)_j \in \mathbb{C}, j = 1,2,3$. The relative maximal Fourier error is defined as*

$$\Delta_{F,\infty}^{(i)} := \frac{\max_j \left| \left(\boldsymbol{f}^{(i)}\right)_j - \left(\boldsymbol{f}_h^{(i)}\right)_j \right|}{\max_j \left| \left(\boldsymbol{f}^{(i)}\right)_j \right|} \in \mathbb{R}_0^+. \tag{II.58}$$

Especially in grating theory and in the beginning of the evolution of FMM (see Sec. III.1.1) only energy conservation was analysed. This means that not the phase relation between diffraction orders, i.e. Fourier coefficients, but their absolute values are compared. This is reasonable in the sense that the periodicity of the geometries investigated with this method already defines the direction of the diffraction orders. Nevertheless, in the context of metrology and in-situ process control of chip design one is much more interested in the correct phase relations. The latter are reflected by $\Delta_F^{(i)}$. With respect to energy conservation and former work we also define the error of the absolute values of the Fourier coefficients.

Definition II.4.9 (Relative absolute Fourier Error)**.** *Let $\boldsymbol{f}^{(i)} \in \mathbb{C}^3$ be the desired Fourier coefficients and $\boldsymbol{f}_h^{(i)} \in \mathbb{C}^3$ their approximation. The relative absolute Fourier error is defined as*

$$\Delta_A^{(i)} := \left| \frac{\left|\boldsymbol{f}^{(i)}\right| - \left|\boldsymbol{f}_h^{(i)}\right|}{\left|\boldsymbol{f}^{(i)}\right|} \right| \in \mathbb{R}_0^+. \tag{II.59}$$

The total relative absolute Fourier error is

$$\Delta_A := \frac{1}{N_f} \sum_i \Delta_A^{(i)}, \tag{II.60}$$

where N_f is the number of diffraction orders of the specific periodic problem.

Again, convergence in energy conservation, i.e. for the Δ_A error, should be carefully compared to the far-field convergence, i.e. Δ_F, itself. From the triangle inequality it follows directly [cf. Eq. (II.51)]

$$\Delta_A \leq \Delta_F. \tag{II.61}$$

Note that analogous to the discussion of Definition II.4.6 one can analyse the contribution of the different diffraction orders to the energy. This lowers the influence of relative errors of the higher diffraction orders on Δ_F and Δ_A by a factor of the cosine of the diffraction angle [cf. Eq. (II.28)].

Chapter III

Numerical Methods

III.1 Fourier Modal Method

The Fourier Modal Method has been a well established method for decades and is known under various synonyms. It was first formulated as the Rigorous Coupled Wave Analysis (RCWA) method. Yet this abbreviation is sometimes referred to as Rigorous Coupled Waveguide Analysis method to emphasize its basics. Commonly used synonyms or variations of the method include the Plane Wave Expansion (PWE), the Eigenmode Expansion (EME) and Eigenmode Expansion Technique (EET) method. Nevertheless, the modern and very often used term for this kind of method is the Fourier Modal Method (FMM). The latter shows both the functional basis of Fourier Transform (FT) and the expansion idea of eigenmodes. That is why throughout this work we usually refer to the numerical method depicted in detail in this section as FMM.

We start by giving a short introduction to the historical evolution of the method from the 1960s to recent work in Section III.1.1. Afterwards, the major theoretical breakthrough of the FMM, the so-called Fourier Factorization Rules (FFR), are motivated, stated and briefly proved. Additionally, we remark on the finite truncation of the infinite problem in Section III.1.3. We end our discussion of the background of the FMM by giving the modern formulations of the common variants in use.

III.1.1 Historical Review

The origins of the FMM date back to the 1960s when Tamir [79] analysed sinusoidal stratified structures in the context of a plane wave description. Yeh [86] focused particularly on the TM computation of these gratings. The founders of modern FMM (or in their terms, RCWA) are Moharam and Gaylord [54]. They generalized the plane wave concept and used RCWA to analyse transmission-grating and reflection-grating behaviour. They compared their method to approximative modal theories that exist at the time and found good and fast convergence of their results. RCWA was formulated as state-space representation, or in modern terms, in the frequency domain.

In 1982 Moharam [55] used a staircase approximation to simulate the diffraction of surface-relief gratings. This approximation of geometry will be further analysed in Section IV.3.2. Furthermore, Moharam and Gaylord published a guide to implement RCWA in a numerically stable way in 1995 [56]. They extended their formulation to TE, TM and conical diffraction and analysed the convergence of the method with respect to the diffraction efficiencies, i.e. to the absolute value of the Fourier coefficients. In the linear plots displayed in their paper the value itself converges well with the number of harmonics. Yet this error estimation appears to be not accurate enough for modern applications since it dismisses phase relations.

The authors expected convergence problems for binary gratings with large periods, deep grooves, TM illumination and conical diffraction - the latter two actually being the same problem. In their opinion RCWA converges to the proper solution and conservation of energy is always satisfied. They emphasized the need to model evanescent modes and their derivation is based on the "coupled-wave equations" for the electric and the magnetic field, derived from Maxwell's equations.

Convergence theory of the FMM had been more or less absent until 1996, when Lalanne [38] and Granet [23] found a solution to the TM convergence problem in numerical experiments. Due to this huge progress in the FMM, Li formulated the famous FFR [43]. He also justified the truncation of the involved Fourier Transform and coined the name FMM for these types of methods.

In 2000 Popov [63] revealed violation of Li's FFR in emerging 3D computations and proposed the decomposition into tangential and normal components of the electric field. In doing so, he invented the so-called Fast Fourier Factorization (FFF). This concept of generalized polarization basis was further developed by Götz [22] and Schuster [75], who automatically generated the normal vector field needed for Popov's FFF. For Photonic Crystals (PhC), Antos [2] proposed a complex polarization basis and called the resulting method complex Fourier Factorization which can be regarded as using elliptic rather than linear polarizations compared to the former ideas.

Finally, in 2010, Essig [18] worked on the so-called Adaptive Spatial Resolution (ASR). This includes the generation of non equally spaced spatial grids, which are not only motivated by the polarization basis as before, but also from the geometric features themselves. This concept was extended to 3D PhCs by Küchenmeister in 2014 [37].

III.1.2 Fourier Factorization Rules

Although the FMM has been used for decades, severe convergence problems had been reported but remained unsolved throughout the literature. In the context of PhCs, Sözüer showed that the FMM shows slow convergence for a dielectric hard-sphere function [78]. In diffraction theory these problems were compensated by higher numerical effort but had to be considered for metallic diffraction gratings [46]. Since these severe problems only occur for TM polarization in 2D, the discontinuity of the electric field at material interfaces was identified as their origin. Both the electric field and the permittivity have discontinuities at these interfaces. The electric displacement field remains continuous.

Lalanne [38] and Granet [23] independently found a numerical solution for the slow convergence by taking the inverse of the Toeplitz matrix of the permittivity function at a specific point in the computation. This procedure is called Inverse Rule or Fast Fourier Factorization [58]. Its mathematical correctness was proven by Li [43] but lacks a physical explanation. Li analysed the local convergence behaviour at the discontinuities pointwise and estimated its error with respect to the number of harmonics. The detailed proof was not accepted by the SIAM Journal of Applied Mathematics and was only published five years later [7].

In the context of Fourier-Galerkin Methods for Photonic Bands Anić recapitulated Li's proof more rigorously and in greater detail [1]. That is why we summarize his findings in Section III.1.2.3 after illustrating the problem with a simple example and defining notation conventions and basic statements.

III.1.2.1 Example

In Section IV.2.1 Fourier Factorization Rules (FFR) are analysed when solving Maxwell's equations. Here, we use a simple example [58, 6] to illustrate the problem of taking the FT of a continuous function which is the product of two discontinuous functions.

Consider the following two periodic functions f and g

$$f(x) = \begin{cases} a & , -\Lambda/2 < x \leq 0 \\ b & , 0 < x \leq -\Lambda/2 \end{cases} \tag{III.1}$$

$$g(x) = \begin{cases} b & , -\Lambda/2 < x \leq 0 \\ a & , 0 < x \leq -\Lambda/2 \end{cases}. \tag{III.2}$$

Their product $h(x) = f(x)g(x) = ab$ is obviously continuous (Fig. III.1). Taking $a = 0.5$ and $b = 2.0$ the 0-th Fourier coefficients are

$$[\![f]\!]_0 = \frac{1}{\Lambda}\int_{-\Lambda/2}^{\Lambda/2} f(x)\,dx = 1.25 \tag{III.3}$$

$$[\![g]\!]_0 = \frac{1}{\Lambda}\int_{-\Lambda/2}^{\Lambda/2} g(x)\,dx = 1.25. \tag{III.4}$$

Calculating the convolution by using only one Fourier harmonic $[\![fg]\!]_0 = [\![f]\!]_0[\![g]\!]_0 = 1.5625$ in a naive way, which will later be called Laurent's Rule, yields a more than 50% error to the exact result $[\![h]\!]_0 = 1.0$. This huge error vanishes already for a truncation to only one harmonic using the so-called Inverse Rule:

$$[\![h]\!]_0 = [\![fg]\!]_0 = \left(\left[\!\left[\frac{1}{f}\right]\!\right]^{-1}\right)_0 [\![g]\!]_0 = \left(\frac{0.5}{2.0} + \frac{0.5}{0.5}\right)^{-1} 1.25 = 1.0. \tag{III.5}$$

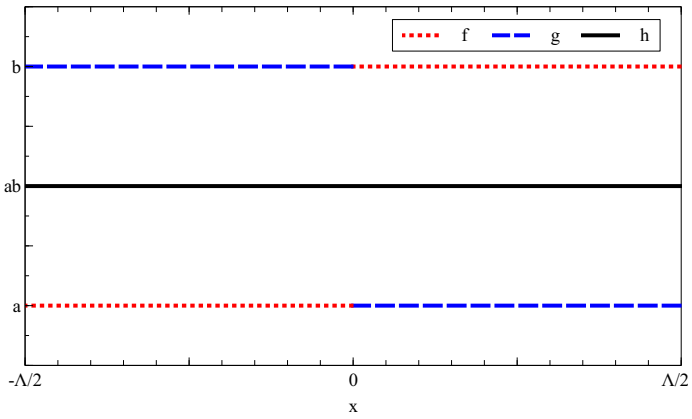


Figure III.1: *Simple example of continuous product $h = fg$ (black solid line) of discontinuous functions f (red dotted line) and g (blue dashed line). Using one harmonic for Fourier convolution with Laurent's Rule yields an error of approximately 56% (cf. main text). When applying the Inverse Rule this error vanishes [Eq. (III.5)].*

III.1.2.2 Notation and Basics

In order to obtain a consistent notation the following definitions are useful. First we define the set of functions of interest $\boldsymbol{P}$:

Definition III.1.1 (Periodic Functions)**.** *Let $\boldsymbol{P}$ be the set of 2π-periodic real valued functions which are piecewise in $C^2([0, 2\pi])$. That means that there exists a $m \in \mathbb{N}$ and an ascending sequence $a_k \in [0, 2\pi]$ such that $0 = a_0 < a_1 < ... < a_m = 2\pi$ and a piecewise representation of $f \in \boldsymbol{P}$ with $f_k \in C^2(a_k, a_{k+1})$ for $k = 0, ..., m-1$.*

Furthermore, we deal with discontinuities. That is why we use U_f to denote the set of discontinuities of f.

Definition III.1.2 (Set of Abscissae of Discontinuities)**.** *Let $f \in \boldsymbol{P}$. The set of abscissae of discontinuities of f is defined as*

$$U_f := \{x_i | f(x_i+) \neq f(x_i-), x_i \in [0, 2\pi]\},$$

where $f(x_i+) := \lim_{x \searrow x_i} f(x)$ and $f(x_i-) := \lim_{x \nearrow x_i} f(x)$.

The so-called *concurrent discontinuities* of f and g are $U_{f,g} := U_f \cap U_g$. Crucial for the following is the fact that f and g yield a continuous product $h = fg$ at the location of concurrent discontinuities. For the electric field $\boldsymbol{E}$ and the permittivity ε this is the case, since the electric displacement field $\boldsymbol{D} = \varepsilon \boldsymbol{E}$ is continuous. In general we call a discontinuity x_i, for which it holds $h(x_i+) = h(x_i-)$, a *complementary discontinuity*. To measure the *size of a discontinuity* we define $f_i^{\models} := f(x_i+) - f(x_i-)$. Since we deal with truncated Fourier series, the following defintions are useful.

Definition III.1.3 (Truncated Laurent Fourier Series)**.** *The truncated Laurent Fourier series of $h = fg$ with $2M + 1$ harmonics is*

$$h^{(M)}(x) := \sum_{n=-M}^{M} h_n^{(M)} \exp(inx)$$

with the truncated Laurent Fourier coefficients

$$h_n^{(M)} := \sum_{m=-M}^{M} f_{n-m} g_m. \tag{III.6}$$

In order to rewrite Eq. (III.6), we define the so-called Toeplitz Matrix $[\![f]\!]$ of f. In doing so, Eq. (III.6) reads as a matrix vector product: $\boldsymbol{h}^{(M)} = [\![f]\!]\boldsymbol{g}$, where $\left(\boldsymbol{h}^{(M)}\right)_n = h_n^{(M)}$ and $(\boldsymbol{g})_n = g_n$.

Definition III.1.4 (Fourier Toeplitz Matrix)**.** *The Toeplitz matrix generated from the Fourier coefficients f_n of f for a convolution with $2M+1$ harmonics is*

$$[\![f]\!] := \begin{pmatrix} f_0 & f_{-1} & f_{-2} & \cdots & \cdots & f_{-2M} \\ f_1 & f_0 & f_{-1} & \ddots & & f_{1-2M} \\ f_2 & f_1 & f_0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & f_{-1} & f_{-2} \\ \vdots & & \ddots & f_1 & f_0 & f_{-1} \\ f_{2M} & \cdots & \cdots & f_2 & f_1 & f_0 \end{pmatrix}. \tag{III.7}$$

The improvement of the FMM is based on the Inverse Rule. We call the reconstruction from this rule the Inverse Fourier Series and define its truncated form:

Definition III.1.5 (Truncated Inverse Fourier Series)**.** *The truncated Inverse Fourier series of $h = fg$ with $2M+1$ harmonics is*

$$\tilde{h}^{(M)}(x) := \sum_{n=-M}^{M} \tilde{h}_n^{(M)} \exp(inx)$$

with the truncated Inverse Fourier coefficients

$$\tilde{h}_n^{(M)} := \sum_{m=-M}^{M} \left(\left[\!\left[\frac{1}{f} \right]\!\right]^{-1} \right)_{nm} g_m,$$

where $[\![1/f]\!]$ is the Fourier Toeplitz Matrix of $1/f$.

To compare the errors introduced by the usual truncated Laurent Fourier series $h^{(M)}(x)$ (Def. III.1.3) and the truncated Inverse Fourier series $\tilde{h}^{(M)}(x)$ (Def. III.1.5), we write the exact reconstruction of the product function h with $2M+1$ harmonics $h_M(x)$.

Definition III.1.6 (Exact truncated Fourier Series)**.** *The exact truncated Fourier series of $h = fg$ with $2M+1$ harmonics is*

$$h_M(x) := \sum_{n=-M}^{M} h_n \exp(inx)$$

with the exact Fourier coefficients

$$h_n := \sum_{m=-\infty}^{\infty} f_{n-m} g_m.$$

It is a well-known result of the convolution of Fourier series that the exact truncated Fourier series $h_M(x)$ corresponds to the original function $h(x)$ in the limit $M \to \infty$ [91]. This means

$$h(x) = \lim_{N\to\infty} \sum_{n=-N}^{N} \left(\lim_{M\to\infty} \sum_{m=-M}^{M} f_{n-m} g_m \exp(inx) \right). \tag{III.8}$$

This section deals with the question in which cases the symmetrically truncated Laurent Fourier series converges to the original function

$$h(x) \overset{?}{\Leftrightarrow} \lim_{M\to\infty} \sum_{n=-M}^{M} \left(\sum_{m=-M}^{M} f_{n-m} g_m \exp(inx) \right) \tag{III.9}$$

and what can be done to speed up the convergence. However, it should be noted that the truncation itself needs to be justified (see Sec. III.1.3). The most important tool for understanding convergence of Fourier convolutions is the following estimation [1].

Theorem III.1.1 (Decay of Fourier Coefficients)**.** *Let $f \in \boldsymbol{P}$ and f be continuous, i.e. $U_f = \varnothing$. Then it holds*

$$|f_n| \leq \frac{C_f}{n^2}$$

with C_f dependent on $\|f_k'\|_\infty$ and $\|f_k''\|_{L^2}$.

III.1.2.3 Theorems of Fourier Factorization

Here, the original theorems of Fourier Factorization [43] are stated and sketches of their proofs [1] are shown. The basic idea of the Inverse Rule is shown in the proof of Theorem III.1.4. Firstly, it holds for the convergence of functions having no concurrent discontinuities:

Theorem III.1.2 (Convergence of Laurent's Rule)**.** *Let $f, g \in \boldsymbol{P}$, $h = fg$ and f and g have no concurrent discontinuities, i.e. $U_{f,g} = \varnothing$. Then the truncated Laurent Fourier series $h^{(M)}(x)$ converges, i.e.*

$$\lim_{M\to\infty} h^{(M)}(x) = h(x)$$

Secondly, an estimation of the convergence of functions showing concurrent disontinuities is given.

Theorem III.1.3 (Convergence of truncated Laurent Fourier Coefficient)**.** *Let $f, g \in \boldsymbol{P}$, $h = fg$ and f and g have concurrent discontinuities, i.e. $U_{f,g} \neq \varnothing$. Then the truncated Laurent Fourier series $h^{(M)}(x)$ has the following error behaviour*

$$h^{(M)}(x) = h_M(x) - \sum_{x_i \in U_{f,g}} \frac{f_i^{\models} g_i^{\models}}{2\pi^2} \Phi_M(x - x_i) - o(1),$$

where $o(1)$ uniformly tends to zero for $M \to \infty$ and

$$\Phi_M(x) := \sum_{n=1}^{M} \frac{\cos(nz)}{n} \sum_{|m|>M} \frac{1}{m-n}.$$

It holds that

$$\lim_{M\to\infty} \Phi_M(x) = \begin{cases} 0 & , x \neq 0 \\ \frac{\pi^2}{4} & , x = 0 \end{cases}.$$

Sketch of the Proof of Theorem III.1.2 and III.1.3.

(i) The proof of the convergence of the truncated Laurent Fourier coefficients is based on a decomposition of f and g into a continuous part $\tilde{f}, \tilde{g}$ and discontinuous parts for each discontinuity $x_i \in U_f, x_j \in U_g$. This idea is based on Theorem F of [29] which itself goes back to [11]. The decomposition used here is

$$f(x) = \tilde{f}(x) + \sum_{x_i \in U_f} \frac{f_i^{\models}}{\pi} \phi(x - x_i)$$

with the periodically extended function $\phi(x) := \frac{1}{2}(\pi - x)$ for $x \in (0, 2\pi)$.

(ii) Then we can rewrite $h = fg$ as

$$h(x) = Q(x) + \frac{1}{\pi} \sum_{x_i \in U_f} f_i^{\models} R(x; x_i) + \frac{1}{\pi} \sum_{x_j \in U_g} g_j^{\models} S(x; x_j) + \frac{1}{\pi^2} \sum_{x_i \in U_f, x_j \in U_g} f_i^{\models} g_j^{\models} T(x; x_i; x_j),$$

where

$$\begin{aligned} Q(x) &:= \tilde{f}(x)\tilde{g}(x) \\ R(x; x_i) &:= \phi(x - x_i)\tilde{g}(x) \\ S(x; x_i) &:= \tilde{f}(x)\phi(x - x_i) \\ T(x; x_i; x_j) &:= \phi(x - x_i)\phi(x - x_j). \end{aligned}$$

(iii) With Theorem III.1.1 and basic estimations one can show that

$$|Q^{(M)}(x) - Q_M(x)| \leq \mathcal{O}\left(\frac{1}{M}\right). \tag{III.10}$$

(iv) With the help of an integral test of series convergence it follows

$$|R^{(M)}(x; x_i) - R_M(x; x_i)| \leq \mathcal{O}\left(\frac{\ln(M)}{M}\right). \tag{III.11}$$

(v) The same integral test as for (III.11) leads to

$$|S^{(M)}(x; x_i) - S_M(x; x_i)| \leq \mathcal{O}\left(\frac{\ln(M)}{M}\right). \tag{III.12}$$

(vi) Lengthy calculations show that for $M \to \infty$

$$T^{(M)}(x; x_i; x_j) - T_M(x; x_i; x_j) = -\frac{1}{2}\Phi_M(x - x_i, x_i - x_j) + o(1), \tag{III.13}$$

where

$$\Phi_M(x_1, x_2) := \frac{1}{2} \sum_{0<|n|\leq M} \frac{\exp(inx_1)}{n} \sum_{|m|>M} \frac{\exp(imx_2)}{m-n}.$$

(vii) For $x_2 := x_i - x_j \neq 0$, which means there are no concurrent discontinuities, and arbitrary $x_1 := x - x_i$ one can show that

$$|\Phi_M(x_1, x_2)| \leq \mathcal{O}\left(\frac{\ln(M)}{M}\right).$$

This proves Theorem III.1.2.

(viii) For $x_2 = 0$, which means there are concurrent discontinuities, and $x_1 \neq 0$ one can show that

$$|\Phi_M(x_1, 0)| \leq \mathcal{O}\left(\frac{\ln(M)}{M}\right)$$

and for $x_1 := x - x_i = 0$, which means at the concurrent discontinuity, it follows

$$\lim_{M\to\infty} |\Phi_M(0, 0)| = \frac{\pi^2}{4}.$$

This proves Theorem III.1.3.

□

We state the central theorem about products of discontinuous functions next. It shows that the truncated Inverse Fourier series (Def. III.1.5) converges to the infinite Fourier series. The basic idea of this rule is shown in (ii) and (iii) of the following proof.

Theorem III.1.4 (Convergence of Inverse Rule)**.** *Let $f, g \in \boldsymbol{P}$, $h = fg$ and the discontinuities of f and g be complementary, i.e. h is continuous. Additionally, let $f(x) \neq 0$ for all $x \in [0, 2\pi)$. If f satisfies either one of the following conditions*

(a) $\mathrm{Re}\,(1/f)$ *does not change sign in* $[0, 2\pi)$ *and* $\inf_{x\in[0,2\pi)} |\mathrm{Re}\,(1/f(x))| > 0$

(b) $\mathrm{Im}\,(1/f)$ *does not change sign in* $[0, 2\pi)$ *and* $\inf_{x\in[0,2\pi)} |\mathrm{Im}\,(1/f(x))| > 0$,

then the truncated Inverse Fourier series $\tilde{h}^{(M)}(x)$ converges, i.e.

$$\lim_{M\to\infty} \tilde{h}^{(M)}(x) = h(x).$$

Sketch of the Proof of Theorem III.1.4.

(i) One can show that

$$\max_{|n|\leq M} \sum_{m=-M}^{M} \left| \left(\left[\!\left[\frac{1}{f} \right]\!\right]^{-1} \right)_{nm} \right| \leq \mathcal{O}(\sqrt{M}). \tag{III.14}$$

This is done with the help of the Cauchy-Schwarz inequality, using the fact that $[\![1/f]\!]$ is a Fourier Toeplitz matrix and one of the conditions on $\mathrm{Re}\,(1/f)$ or $\mathrm{Im}\,(1/f)$, respectively. Additionally, one uses $\|B\|_\infty \leq \sqrt{n}\|B\|_2$ for any $B \in \mathbb{C}^{n\times n}$.

(ii) It obviously holds $g = 1/f$ h. We use the estimations (III.10)-(III.13) of the proof of Theorem III.1.3 and the fact that h is continuous to conclude for this truncated Laurent Fourier coefficient

$$g_n = \sum_{n=-M}^{M} g_n^{(M)} - \delta_n = \sum_{n=-M}^{M} \left(\frac{1}{f}\right)_{n-m} h_m - \delta_n = \sum_{n=-M}^{M} \left[\!\left[\frac{1}{f}\right]\!\right]_{nm} h_m - \delta_n, \tag{III.15}$$

where $\delta_n = \mathcal{O}\left(\ln(M)/M^2\right)$ is determined by the derivation of (III.11). All other terms decay faster or are zero, since h is continuous.

(iii) Now we use the inverse of (III.15)

$$h_n = \sum_{n=-M}^{M} \left(\left[\!\left[\frac{1}{f} \right]\!\right]^{-1} \right)_{nm} (g_m + \delta_m)$$

to obtain

$$\begin{aligned} h_n - \tilde{h}_n^{(M)} &= h_n - \sum_{m=-M}^{M} \left(\left[\!\left[\frac{1}{f} \right]\!\right]^{-1} \right)_{nm} g_m \\ &= \sum_{n=-M}^{M} \left(\left[\!\left[\frac{1}{f} \right]\!\right]^{-1} \right)_{nm} (g_m + \delta_m) - \sum_{m=-M}^{M} \left(\left[\!\left[\frac{1}{f} \right]\!\right]^{-1} \right)_{nm} g_m \\ &= \sum_{m=-M}^{M} \left(\left[\!\left[\frac{1}{f} \right]\!\right]^{-1} \right)_{nm} \delta_m. \end{aligned}$$

(iv) Together with (III.14) this leads to

$$\left| \tilde{h}^{(M)}(x) - h_M(x) \right| \leq \mathcal{O}\left(\frac{\ln(M)}{\sqrt{M}} \right)$$

and proves Theorem III.1.4.

□

III.1.2.4 Interpretation of Inverse Rule

The mathematical procedure of taking the inverse Toeplitz Matrix in (iii) in the proof of Theorem III.1.4 still lacks a physical explanation. Banerjee and Jarem [6] tried to find this explanation with the help of the example in Section III.1.2.1. They stated that the truncated Fourier series of f and g are $f^{(M)}(0) = g^{(M)}(0) = (a+b)/2$ at the discontinuity. Using the truncated Laurent Fourier series one obtains for the product $h^{(M)}(0) = (a+b)^2/4 \neq h(0)$.

Their reformulation of the Inverse Rule consists of the following steps: (a) Take the reconstruction of the truncated Fourier Transform of $1/f(x)$ giving $f^{REC}(x)$. (b) Invert $f^{REC}(x)$ leading to $1/f^{REC}(x)$. (c) Take the Fourier Transform of $1/f^{REC}(x)$ and build the product with the truncated Fourier Transform of g. Following this procedure, they obtain $\bar{h}^{(M)}(0) = ab = h(0)$. This behaviour is explained with non-zero derivatives of the truncated Laurent Fourier series $h^{(M)}$ compared to Banerjee's and Jarem's constant Inverse Rule series $\bar{h}^{(M)}$, which has zero derivatives.

Although the procedure above initially reads like the Inverse Rule, Li pointed out that it is not [45]. Banerjee and Jarem used a rule which does the multiplication in real space rather than in

Fourier space as the proper Inverse Rule. However, the latter is not a simple convolution in Fourier space either since the inverse of the truncated Fourier Toeplitz matrix is generally not a Fourier Toeplitz matrix. To clarify the comparison to the procedure above, we summarize the Inverse Rule as follows: (a) Take the truncated Fourier transform of $1/f(x)$ and obtain the corresponding Fourier Toeplitz matrix $[\![1/f]\!]$. (b) Invert the Fourier Toeplitz matrix $[\![1/f]\!]^{-1}$ and build the product with the truncated Fourier Transform of g.

Furthermore, the basic idea of Definition III.1.5 is motivated by a well-known result from theory of Toeplitz matrices. Let $[\![f]\!]^{(\infty)}$ be the infinite version of $[\![f]\!]$ (Def. III.1.4), i.e. the Fourier Toeplitz matrix for $M \to \infty$. Then it holds [83]

$$[\![f]\!]^{(\infty)} = \left(\left[\!\left[\frac{1}{f} \right]\!\right]^{(\infty)} \right)^{-1}. \tag{III.16}$$

The rather complicated conditions of Li's Theorem III.1.4 on the Inverse Rule correspond to the one in Theorem 1 of [83]: a necessary and sufficient condition for (III.16) is that $1/f$ is essentially bounded. This means that there exists a constant $C < \infty$ such that $\{x : |1/f(x)| > C\}$ has zero measure. This again corresponds to the inf-criteria of Li's Theorem.

However, for the convergence improvement of the Inverse Rule we need the additional condition of no sign change of $\mathrm{Re}\,(1/f)$ or $\mathrm{Im}\,(1/f)$. They seem to be the source of the convergence improvement for functions $h = fg$ with complementary discontinuities of f and g and limit the efficient application of the FMM in TM polarization to non-metallic structures, i.e. there should not be a sign change in $\mathrm{Re}\,(\varepsilon)$ [typically dielectrics are involved, so condition (b) of Li's Theorem III.1.4 does already not hold].

III.1.3 Matrix Truncation

In the derivation of the FMM, Maxwell's equations (II.1)-(II.4) are expanded in an infinite Fourier series. However, the corresponding eigenvalue problem (II.18) is solved for a finite number of Fourier harmonics. This truncation is called a reduction method. Often, the convergence of the solution of the truncated problem to the solution of the infinite system is simply assumed. Nevertheless, for eigenvalue problems, there are examples for which this is not true: Sayer constructed an infinite system with entries made out of Legendre polynomials [72]. For this system the solution obtained with the reduction method does not converge to the eigenvalues of the infinite problem. That is why we follow the arguments in [44] and [7] respectively, to show that for the FMM in TE polarization the truncation is partly justified rigorously.

The so-called classical theory of determinants of infinite order is used in [44]. It deals with the convergence of the determinant of the finite problem to the one of the infinite problem. In order to assure convergence the following theorem by Poincaré formulates a condition.

Theorem III.1.5 (Infinite Eigenvalue Problem I)**.** *Let $A = \{A_{ik} = \delta_{ik} + a_{ik}\}$ be a matrix of infinite order. For the determinant of A to be absolutely convergent, it is sufficient that*

$$\sum_{i,k} |a_{ik}| < \infty. \tag{III.17}$$

For discontinuous permittivity profiles this theorem is not sufficient, since the condition (III.17) is not fulfilled (the harmonic series is divergent). That is why we replace (III.17) with the help of the following theorem.

Theorem III.1.6 (Infinite Eigenvalue Problem II)**.** *Let $A = \{A_{ik} = \delta_{ik} + a_{ik}\}$ be a matrix of infinite order. For the determinant of A to be absolutely convergent, it is sufficient that $\sum_i |a_{ii}| < \infty$ and that*

$$\sum_{i,k} \left| a_{ik} \frac{x_i}{x_k} \right|^2 < \infty, \tag{III.18}$$

where $\{x_i\}$ is a sequence of nonzero numbers.

For simplicity we rewrite the eigenvalue problem (II.18) for TE polarization [7]:

$$(\rho + \alpha_n^2) E_{z,n} = k_0^2 \mu \sum_m \varepsilon_{n-m} E_{z,m} \tag{III.19}$$

$$\sum_m \left(\delta_{nm} + \frac{\widetilde{\varepsilon}_{n-m}}{k_0^2 \mu \varepsilon_0 - \alpha_n^2 - \rho} \right) E_{z,m} = 0 \tag{III.20}$$

with the eigenvalues ρ and $\widetilde{\varepsilon}_{n-m} = k_0^2 \mu \varepsilon_{n-m}$ for $n \neq m$ and $\widetilde{\varepsilon}_0 = 0$, $\alpha_n = \alpha_0 + nK$ with $\alpha_0 = k_0 \sqrt{\varepsilon\mu} \sin(\theta_i)$ and the grating vector K. For continuous ε the convergence of the truncated system (III.20) is guaranteed with Theorem III.1.5, since $\widetilde{\varepsilon} \leq O(1/n^2)$. For discontinuous permittivity profiles, however, $\tilde{\varepsilon} = O(1/n)$.

Then Theorem III.1.6 ensures convergence of the determinants: We define $\Omega(R, r)$ to be the disk of radius R in the complex plane centred at the origin, excluding the disk of radius r centred at $k_0^2 \mu \varepsilon_0 - \alpha_n^2$. For $\{x_i\}$ we choose $x_i = i$ for $i \neq 0$ and $x_0 = 1$. It holds that $\sum_{i,k} |a_{i,k} x_i / x_k|$ is uniformly convergent in $\Omega(R, r)$. This together with estimations of the error bounds of the truncated determinants [7] proves the convergence of a major system of linear equations of the FMM in TE polarization.

Nevertheless, it should be noted that $\Omega(R, r)$ does not include the formerly described discs of radius r. This means that the solution of (III.20) converges non-uniformly over the whole complex plane. The excluded points are supposed to determine the overall convergence of the FMM. This notion corresponds to the matter of non-uniform pointwise-convergence of the truncated convolution described in the previous section. Furthermore, to the best of our knowledge, there is no proof of the justification of matrix truncation for TM polarization, i.e. $E_{z,n} = O(1/n)$. In addition, the convergence of (III.20) does not include the convergence of the full FMM, since an additional eigenvalue problem (III.34) has to be solved, which will be described in the formulation of the FMM in the next section.

III.1.4 FMM Formulations

III.1.4.1 Basic Eigenvalue Problem

In this section we state the formulations of the different variants of the FMM. We follow the derivation of [49] and adapt and unify notation slightly. First of all, due to Eq. (II.17) we write the magnetic field as a pseudo-Fourier series

$$\boldsymbol{H}(\boldsymbol{r}_\perp, z) = \sum_{\boldsymbol{G}_m} \boldsymbol{H}_{\boldsymbol{G}_m}(z) \, \mathrm{e}^{i(\boldsymbol{k}_\perp + \boldsymbol{G}_m) \cdot \boldsymbol{r}_\perp}. \tag{III.21}$$

Since we assume layers that are uniform in the z-direction (see Sec. II.2.2) the expansion is done for $\boldsymbol{r}_\perp, \boldsymbol{k}_\perp$ in the xy-plane. Compared to the previous sections, we changed the FT from a 1D series to 2D with the help of reciprocal lattice vectors $\boldsymbol{G}_m$. It is not straightforward to choose a finite number of these reciprocal lattice vectors in the two-dimensional lattice to obtain a finite dimensional basis set for the FMM. However, due to symmetry arguments for the Toeplitz matrix (Def. III.1.4), for every $\boldsymbol{G}_m$ we also use the vector $-\boldsymbol{G}_m$. This ensures convergence of the convolution analysed in the previous sections.

Furthermore, for general lattices the so-called circular truncation is best suited, since it represents a trade-off between using high diffraction orders and limit the total number of diffraction orders by using M reciprocal lattice vectors which lie inside a circle of the reciprocal lattice [49]. That is why we change the meaning of M to $M := |\{\boldsymbol{G}_m\}|$, which depends on the truncation strategy.

We define the vector $\underline{\boldsymbol{h}}(z) = (\boldsymbol{H}_{\boldsymbol{G}_1}(z), \boldsymbol{H}_{\boldsymbol{G}_2}(z), ...)^T \in (\mathbb{C}^3)^M$ of the three-dimensional Fourier coefficients and similarly $\underline{\boldsymbol{e}}(z)$. Using these Fourier expansions in the time-harmonic Maxwell's equations (II.1) and (II.2) for current-free media ($\boldsymbol{\mathcal{J}} = 0$), we obtain:

$$i\underline{K}_y \underline{h}_z(z) - \partial_z \underline{h}_y(z) = -i\omega \underline{d}_x(z) \tag{III.22}$$

$$\partial_z \underline{h}_x(z) - i\underline{K}_x \underline{h}_z(z) = -i\omega \underline{d}_y(z) \tag{III.23}$$

$$i\underline{K}_x \underline{h}_y(z) - i\underline{K}_y \underline{h}_x(z) = -i\omega \underline{d}_z(z) \tag{III.24}$$

$$i\underline{K}_y \underline{e}_z(z) - \partial_z \underline{e}_y(z) = i\omega \underline{h}_x(z) \tag{III.25}$$

$$\partial_z \underline{e}_x(z) - i\underline{K}_x \underline{e}_z(z) = i\omega \underline{h}_y(z) \tag{III.26}$$

$$i\underline{K}_x \underline{e}_y(z) - i\underline{K}_y \underline{e}_x(z) = i\omega \underline{h}_z(z), \tag{III.27}$$

where $\left(\underline{K}_{x,y}\right)_{nm} := \delta_{mn}\left[\boldsymbol{k}_{x,y} + \left(\boldsymbol{G}_m\right)_{x,y}\right] \in \mathbb{C}^{M\times M}$ is a diagonal matrix of diffraction wavenumbers in the x- and the y-direction, respectively. These are determined by the geometry, i.e. the reciprocal lattice vectors $\boldsymbol{G}_m$, and the incoming wave vector $\boldsymbol{k}$.

In order to link the equations for the electric field to those for the electric displacement field, we need the following relation

$$\begin{pmatrix} -\underline{d}_y(z) \\ \underline{d}_x(z) \\ \underline{d}_z(z) \end{pmatrix} = \begin{pmatrix} \underline{\mathcal{E}} & 0 \\ 0 \quad 0 & [\![\varepsilon]\!] \end{pmatrix} \begin{pmatrix} -\underline{e}_y(z) \\ \underline{e}_x(z) \\ \underline{e}_z(z) \end{pmatrix}, \tag{III.28}$$

where the Toeplitz matrix is of the form $[\![\varepsilon]\!] := \{\varepsilon_{\boldsymbol{G}_m-\boldsymbol{G}_n}\} \in \mathbb{C}^{M\times M}$. Here, the Fourier coefficients of the reciprocal lattice vectors are $\varepsilon_{\boldsymbol{G}} = 1/|\Psi| \int_\Psi \varepsilon(\boldsymbol{r}_\perp)\exp(i\boldsymbol{G}\cdot\boldsymbol{r}_\perp)d\boldsymbol{r}_\perp$ with the unit cell Ψ in the xy-plane. Due to the uniformity in the z-direction, the z-component of the electric field is always tangential to material interfaces. Therefore, Laurent's Rule can be used for the last component of (III.28) (see Sec. III.1.2.3). For the x- and y-component proper Fourier Factorization Rules should be used as explained in the previous sections. This choice is subject to the different variants of the FMM. So $\underline{\mathcal{E}} \in \left(\mathbb{C}^{M\times M}\right)^{2\times 2}$ depends on the formulation of the FMM and will be analysed in detail in the next section.

Using equations (III.22), (III.23) and (III.27), we obtain in matrix notation

$$\left(\omega^2\underline{\mathcal{I}} - \underline{\mathcal{K}}\right)\begin{pmatrix} \underline{h}_x(z) \\ \underline{h}_y(z) \end{pmatrix} = -i\omega\partial_z \begin{pmatrix} -\underline{e}_y(z) \\ \underline{e}_x(z) \end{pmatrix} \quad \text{with} \tag{III.29}$$

$$\underline{\mathcal{K}} := \begin{pmatrix} \underline{K}_y[\![\varepsilon]\!]^{-1}\underline{K}_y & -\underline{K}_y[\![\varepsilon]\!]^{-1}\underline{K}_x \\ -\underline{K}_x[\![\varepsilon]\!]^{-1}\underline{K}_y & \underline{K}_x[\![\varepsilon]\!]^{-1}\underline{K}_x \end{pmatrix}, \tag{III.30}$$

where $\underline{\mathcal{K}}, \underline{\mathcal{I}} \in \left(\mathbb{C}^{M\times M}\right)^{2\times 2}$ and $\underline{\mathcal{I}}$ is the identity matrix. Eliminating the z-components of (III.24), (III.25) and (III.26) as well, yields

$$\left(\omega^2\underline{\mathcal{E}} - \underline{\mathfrak{K}}\right)\begin{pmatrix} -\underline{e}_y(z) \\ \underline{e}_x(z) \end{pmatrix} = -i\omega\partial_z \begin{pmatrix} \underline{h}_x(z) \\ \underline{h}_y(z) \end{pmatrix} \quad \text{with} \tag{III.31}$$

$$\underline{\mathfrak{K}} := \begin{pmatrix} \underline{K}_x^2 & \underline{K}_x\underline{K}_y \\ \underline{K}_y\underline{K}_x & \underline{K}_y^2 \end{pmatrix}, \tag{III.32}$$

where $\underline{\mathfrak{K}} \in \left(\mathbb{C}^{M\times M}\right)^{2\times 2}$.

For layers which are uniform in the z-direction we expand the electromagnetic fields into eigenmodes with a simple $\exp(i\beta_n z)$ dependence (see Sec. II.2.2). Due to Maxwell's equation (II.3), we use the following form for the n-th eigenmode:

$$\boldsymbol{H}_n(z) = \sum_{\boldsymbol{G}_m}\left[\phi^{(x)}_{\boldsymbol{G}_m}\hat{\boldsymbol{x}} + \phi^{(y)}_{\boldsymbol{G}_m}\hat{\boldsymbol{y}} - \frac{\left(\underline{K}_x\right)_{mm}\phi^{(x)}_{\boldsymbol{G}_m} + \left(\underline{K}_x\right)_{mm}\phi^{(y)}_{\boldsymbol{G}_m}}{\beta_n}\hat{\boldsymbol{z}}\right] \mathrm{e}^{i(\boldsymbol{k}_\perp+\boldsymbol{G}_m)\cdot\boldsymbol{r}_\perp + i\beta_n z}, \tag{III.33}$$

where $\phi^{(x,y)}_{\boldsymbol{G}_m} \in \mathbb{C}$ are expansion coefficients. The vector of Fourier coefficients, which was defined before, has now the form $\underline{\boldsymbol{h}}(z) = \left[\boldsymbol{\phi}^{(x)}\hat{\boldsymbol{x}} + \boldsymbol{\phi}^{(y)}\hat{\boldsymbol{y}} - \beta_n^{-1}(\underline{K}_x\boldsymbol{\phi}^{(x)} + \underline{K}_y\boldsymbol{\phi}^{(y)})\hat{\boldsymbol{z}}\right]\exp(i\beta_n z)$, where $\boldsymbol{\phi}^{(x,y)} := \{\phi^{(x,y)}_{\boldsymbol{G}_m}\} \in \mathbb{C}^M$.

Now, we insert the expansion (III.33) into (III.29) and (III.31). Additionally, we use $\underline{\mathcal{K}}\,\underline{\mathfrak{K}} = 0$ to obtain the eigenvalue problem for the n-th eigenmode:

$$\left[\underline{\mathcal{E}}\left(\omega^2\underline{\mathcal{I}} - \underline{\mathcal{K}}\right) - \underline{\mathfrak{K}}\right]\underline{\Phi}_n = \beta_n^2\underline{\Phi}_n, \tag{III.34}$$

where $\underline{\Phi}_n := (\boldsymbol{\phi}^{(x)}_n, \boldsymbol{\phi}^{(y)}_n)^T \in (\mathbb{C}^M)^2$ with the Fourier coefficients $\boldsymbol{\phi}^{(x,y)}_n$. Equation (III.34) is the core equation of the FMM which is solved to compute the eigenmodes of the layered media. Its size obviously scales with the number of reciprocal lattice vectors $\boldsymbol{G}$. So a geometry consisting of N layers has a storage requirement of $O(M^2N)$. The eigenvalue problem is solved in S^4 [70] (the software package used for FMM simulations in this work) using standard QR algorithms which require $O(M^3)$ operations. That is why the total simulation time is $O(NM^3)$.

After having solved the eigenvalue problem (III.34), the fields are expanded into forward $[\exp(i\beta_n z)]$

and backward [$\exp(-i\beta_n z)$] propagating modes (see Sec. II.2.2 and [49]). Subsequently, the S-matrices of the layers are computed and an incoming field is propagated through the whole layered structure using the S-matrix algorithm as noted before.

III.1.4.2 Variants of the FMM

In the following, we describe the different variants of the FMM. On the one hand, these break down into two different ways of computing Fourier coefficients: either using a closed-form FT or the discretized Fast Fourier Transform. On the other hand, the different formulations are related to applying the correct Fourier Factorization Rules in the xy-plane, which yields different matrices $\underline{\mathcal{E}}$ of the eigenvalue problem (III.34).

Closed-form Fourier Transform In nano-optics many geometries can be described by extrusions of 2D shapes. These shapes might include rectangles, circles, ellipses and simple closed polygons, whereas the latter can be regarded as the generalized form of all shapes. That is why the layering algorithm described in the next section uses polygons to generally describe any geometry. The shapes are specified by a constant permittivity ε. We need to obtain the Fourier coefficients of the Toeplitz matrix $[\![\varepsilon]\!]$ and the general relation between the electric field and the electric displacement field $\underline{\mathcal{E}}$, respectively. For these we need to compute integrals $f_{\boldsymbol{G}} = 1/|\Psi| \int_\Psi f(\boldsymbol{r}_\perp)\exp(i\boldsymbol{G}\cdot\boldsymbol{r}_\perp)d\boldsymbol{r}_\perp$ over the unit cell Ψ for functions f which are related to the permittivity ε (cf. next paragraphs for details). For all shapes mentioned above closed-form Fourier transforms of their indicator functions exist [52], so the exact Fourier coefficients $f_{\boldsymbol{G}}$ can be computed.

Fast Fourier Transform A standard technique to compute the discretized FT is the Fast Fourier Transform (FFT) [65]. The permittivity is discretized onto a grid and the Fourier coefficients are approximated using the FFT. To increase accuracy of the FFT an oversampling factor can be applied: the spatial grid is more finely discretized than the desired frequency discretization for the Fourier coefficients. Nevertheless, in 2D, the FFT is subject to an inherent staircasing effect in the xy-plane. Although this can be avoided by calculating exact pixel overlap [49], the closed-form FT is more accurate and does not yield slower simulation times compared to the FFT.

However, the recently developed approach of Adaptive Spatial Resolution (ASR) [18] uses an adaptive spatial grid. After applying Gaussian smoothing of the specified layer geometry, a spatial grid is adapted to the specific geometry by minimizing an energy functional. This avoids the 2D staircasing effect and reduces the numerical effort of simple oversampling for the FFT, since only regions of varying ε are finely discretized. This is currently not implemented in the software package S^4 used for the study at hand and will not be analysed further.

Subpixel Averaging Improving standard FFT can be done by using subpixel averaging. Here, for each discretization pixel an anisotropic permittivity tensor is computed [20]. Although this is a standard technique in FDTD simulations, we do not find improvements in speed and accuracy for the FMM (see Sec. IV). Accordingly, we do not further analyse the theoretical background of subpixel averaging.

Fourier Factorization Rules Section III.1.2 dealt with the mathematical justification of using different convolutions in Fourier space for the electric displacement field $\boldsymbol{D} = \varepsilon\boldsymbol{E}$ for normal and tangential components of $\boldsymbol{E}$. This was due to the conditions on continuity of $\boldsymbol{D}$. We found that for tangential components (TE polarization) the simple Laurent's Rule (Def. III.1.3) can be used for the Fourier coefficients of ε, while for normal components (TM polarization) the Inverse Rule (Def. III.1.5) is needed for fast convergence. Including these findings in the formalism developed above, for 2D problems (1D periodicity in the x-direction) the following matrix is used to relate the in-plane components of $\boldsymbol{D}$ and $\boldsymbol{E}$:

$$\underline{\mathcal{E}} = \begin{pmatrix} [\![\varepsilon]\!] & 0 \\ 0 & [\![1/\varepsilon]\!]^{-1} \end{pmatrix}. \tag{III.35}$$

Normal Vector Method In order to generalize the use of proper Fourier Factorization Rules for 3D problems (two-fold periodicity in the x- and the y-direction), vector fields are applied to decompose the in-plane components of E_x and E_y into tangential (E_t) and normal (E_n) parts. That is why a smooth vector field $\boldsymbol{t} = (t_x, t_y)^T$, which is tangential to all material interfaces, is applied in

the xy-plane. In doing so, we obtain

$$\begin{pmatrix} E_t \\ E_n \end{pmatrix} = \begin{pmatrix} t_x & -t_y^* \\ t_y & t_x^* \end{pmatrix}^{-1} \begin{pmatrix} E_x \\ E_y \end{pmatrix}. \tag{III.36}$$

Now we simplify

$$\begin{pmatrix} -D_y \\ D_x \end{pmatrix} = T \begin{pmatrix} \varepsilon & 0 \\ 0 & (1/\varepsilon)^{-1} \end{pmatrix} T^{-1} \begin{pmatrix} -E_y \\ E_x \end{pmatrix}, \text{ where } T := \begin{pmatrix} t_y & t_x^* \\ -t_x & t_y^* \end{pmatrix}, \tag{III.37}$$

by using the standard inversion formula for 2×2 matrices for T^{-1}. Furthermore, we denote $\Delta = \varepsilon - (1/\varepsilon)^{-1}$:

$$\begin{pmatrix} -D_y \\ D_x \end{pmatrix} = \left[\begin{pmatrix} \varepsilon & 0 \\ 0 & \varepsilon \end{pmatrix} - \begin{pmatrix} \Delta & 0 \\ 0 & \Delta \end{pmatrix} P \right] \begin{pmatrix} -E_y \\ E_x \end{pmatrix}, \tag{III.38}$$

where

$$P := \frac{1}{|t_x|^2 + |t_y|^2} \begin{pmatrix} |t_y|^2 & t_x^* t_y \\ t_x t_y^* & |t_x|^2 \end{pmatrix}. \tag{III.39}$$

By Fourier transforming Eq. (III.38), we obtain

$$\underline{\mathcal{E}} = [\![\varepsilon]\!]\underline{\mathcal{I}} - ([\![\Delta]\!]\,\underline{\mathcal{I}})\,\underline{P}, \tag{III.40}$$

where $[\![\Delta]\!]$ denotes the Toeplitz matrix of Δ and $\underline{P}$ denotes the FT of P. This can be regarded as a correction to the simple Laurent's Rule $\underline{\mathcal{E}} = [\![\varepsilon]\!]\underline{\mathcal{I}}$.

The first application of using a vector field $\boldsymbol{t}$ to decompose the in-plane components of $\boldsymbol{E}$ within the FMM is due to Schuster et al. [75]. It should be noted that these automatically generated vector fields suffer from slow convergence for locations where $\boldsymbol{t}$ vanishes: for these locations the normalization $1/(|t_x|^2 + |t_y|^2)$ is undefined leading to convergence problems of the Fourier Transform $\underline{P}$.

Jones Vector Field Instead of simplifying Eq. (III.37), one can directly Fourier transform this equation, yielding

$$\underline{\mathcal{E}} = ([\![T]\!]\underline{\mathcal{I}}) \begin{pmatrix} [\![\varepsilon]\!] & 0 \\ 0 & [\![(1/\varepsilon)]\!]^{-1} \end{pmatrix} ([\![T^{-1}]\!]\underline{\mathcal{I}}). \tag{III.41}$$

However, instead of using the vector field $\boldsymbol{t}$ here, Antos [2] proposed using a complex polarization basis in order to obtain vector fields which are smoother over the whole unit cell. That is why a Jones vector field $\boldsymbol{J}$ is used instead of $\boldsymbol{t}$ for the derivation of the Toeplitz matrix $[\![T]\!]$ in Eq. (III.41) [cf. $T(t_x, t_y)$ in Eq. (III.37) which becomes $T(J_x, J_y)$]. The Jones vector field is defined pointwise as

$$\boldsymbol{J} = \frac{\mathrm{e}^{i\theta}}{|\boldsymbol{t}|} \begin{pmatrix} t_x & -t_y \\ t_y & t_x \end{pmatrix} \begin{pmatrix} \cos\varphi \\ i\sin\varphi \end{pmatrix}, \tag{III.42}$$

where $\boldsymbol{t}$ is uniformly scaled to have maximal unit length, $\theta = \sphericalangle(\boldsymbol{t})$ and $\varphi = \pi/8(1 + |\boldsymbol{t}|\cos\pi)$.

III.2 Finite Element Method

In this section, we briefly state the basics of the FEM. We follow the derivations of [40, 57, 16]. For details, definitions and rigorous treatments, refer to these descriptions. First, we formulate Maxwell's equations in the so-called variational, or weak, formulation and mention the fundamental vector spaces involved. Afterwards, we mention the finite dimensional discretization in order to obtain an algorithmic suitable problem. Throughout this work we use the software package *JCMsuite* [32] which models the exterior FEM domain with so-called Perfectly Matched Layers (PML), the idea of which we explain in brief. In Section III.2.3, we outline elementary estimations of the convergence of the FEM. Finally, we use an example to illustrate the use of enhancements of simple FEM with the so-called *hp*-adaptivity.

III.2.1 Weak Formulation

In the following, we use a Galerkin method to reformulate Maxwell's equation (II.13) in weak, i.e. integral form. First of all, our goal is to solve Maxwell's equations within an *interior domain* $\Omega \subset \mathbb{R}^3$. It is bounded by $\partial\Omega$ and the so-called *exterior domain* is $\Omega_{\text{ext}} = \mathbb{R}^3/\Omega$. The scattering problem of an incident field $\boldsymbol{E}_i$ is formulated with an outwards propagating scattered field $\boldsymbol{E}_s$. This viewpoint yields the total field $\boldsymbol{E} = \boldsymbol{E}_i + \boldsymbol{E}_s$. Here, we derive the variational formulation for the total field in the interior domain Ω.

Equation (II.13), which is derived from Maxwell's equations, is multiplied with a so-called *test function* $\boldsymbol{v} \in [C^\infty(\Omega)]^3$ and the result is integrated over the interior domain:

$$\int_\Omega d^3\boldsymbol{r} \left[\boldsymbol{v}^* \cdot \left(\boldsymbol{\nabla} \times \frac{1}{\mu}\boldsymbol{\nabla} \times \boldsymbol{E}\right) - \omega^2 \boldsymbol{v}^* \cdot \varepsilon \boldsymbol{E}\right] = 0. \tag{III.43}$$

We integrate by parts and define the sesquilinear form a_{int} as follows,

$$a_{\text{int}}(\boldsymbol{v}, \boldsymbol{E}) - \int_{\partial\Omega} d\boldsymbol{s}\ \boldsymbol{v}^* \times \left(\frac{1}{\mu}\boldsymbol{\nabla} \times \boldsymbol{E}\right) = 0, \text{ where} \tag{III.44}$$

$$a_{\text{int}}(\boldsymbol{v}, \boldsymbol{E}) := \int_\Omega d^3\boldsymbol{r} \left[(\boldsymbol{\nabla} \times \boldsymbol{v}^*) \cdot \frac{1}{\mu}(\boldsymbol{\nabla} \times \boldsymbol{E}) - \omega^2 \boldsymbol{v}^* \varepsilon \boldsymbol{E}\right]. \tag{III.45}$$

Due to (III.44), it holds $\boldsymbol{\nabla} \times \boldsymbol{E} \in \left[L^2(\Omega)\right]^3$. Therefore, we define the so-called Sobolev space

$$H(\text{curl}, \Omega) := \left\{\boldsymbol{v} \in \left[L^2(\Omega)\right]^3 \middle| \boldsymbol{\nabla} \times \boldsymbol{v} \in \left[L^2(\Omega)\right]^3\right\}. \tag{III.46}$$

Furthermore, due to Maxwell's equation (II.12) for charge-free systems ($\rho = 0$), we introduce the Sobolev space

$$H^1(\Omega) := \left\{v \in L^2(\Omega) \middle| \boldsymbol{\nabla} v \in \left[L^2(\Omega)\right]^3\right\} \tag{III.47}$$

and $H_0^1(\Omega)$ as the subset of functions in $H^1(\Omega)$ which have compact support on Ω. For these functions, one can show that $v|_{\partial\Omega} = 0$ for $v \in H_0^1(\Omega)$. We multiply (II.12) with $v \in H_0^1(\Omega)$, integrate by parts and use $v|_{\partial\Omega} = 0$ to obtain

$$\int_\Omega d^3\boldsymbol{r}\ \ (\boldsymbol{\nabla} v)^* \cdot \varepsilon \boldsymbol{E} = 0. \tag{III.48}$$

This guides us to the kernel of the curl operator $H^0(\text{curl}, \Omega) := \{\boldsymbol{\nabla} v | v \in H_0^1(\Omega)\}$ and Eq. (III.48) reads: $\int_\Omega d^3\boldsymbol{r}\ \boldsymbol{v}^* \cdot \varepsilon \boldsymbol{E} = 0\ \forall \boldsymbol{v} \in H^0(\text{curl}, \Omega)$. The last requirement for the weak formulation is the so-called Helmholtz decomposition of $H(\text{curl}, \Omega)$. It reads

$$H(\text{curl}, \Omega) = H^\perp(\text{curl}, \Omega) \oplus H^0(\text{curl}, \Omega), \text{ with} \tag{III.49}$$

$$H^\perp(\text{curl}, \Omega) := \left\{\boldsymbol{v} \in H(\text{curl}, \Omega) \middle| \int_\Omega d^3\boldsymbol{r}\ \boldsymbol{v}^* \cdot \varepsilon \boldsymbol{w} = 0\ \forall \boldsymbol{w} \in H^0(\text{curl}, \Omega)\right\}. \tag{III.50}$$

From Eq. (III.48) it follows $\boldsymbol{E} \in H^\perp(\text{curl}, \Omega)$. Finally, we state the weak formulation of Maxwell's equations in the interior domain:

$$\begin{aligned} &\text{Find } \boldsymbol{E} \in H^\perp(\text{curl}, \Omega) \text{ such that} \\ &a_{\text{int}}(\boldsymbol{v}, \boldsymbol{E}) - \int_{\partial\Omega} d\boldsymbol{s}\ \boldsymbol{v}^* \times \left(\frac{1}{\mu}\boldsymbol{\nabla} \times \boldsymbol{E}\right) = 0\ \forall \boldsymbol{v} \in H^\perp(\text{curl}, \Omega). \end{aligned} \tag{III.51}$$

III.2.2 Discretization and Perfectly Matched Layers

In order to solve the weak formulation (III.51), one chooses a finite-dimensional subspace $W_h \subset H^\perp(\text{curl}, \Omega)$. The parameter h is a discretization parameter. In the classical formulation of the FEM [57], one can think of h as the maximal length of the discretized parts of a mesh of Ω. The variational formulation of the discrete problem is

$$\begin{aligned} &\text{Find } \boldsymbol{E}_h \in W_h \text{ such that} \\ &a_{\text{int}}(\boldsymbol{v}, \boldsymbol{E}_h) = R(\boldsymbol{v})\ \forall \boldsymbol{v} \in W_h, \end{aligned} \tag{III.52}$$

where we substitute $R(\boldsymbol{v})$ for right-hand side terms which will be elaborated in the next paragraphs. Let $\{\boldsymbol{v}_i\}$ be a basis of W_h. We expand the solution in this basis, yielding $\boldsymbol{E}_h = \sum_i u_i \boldsymbol{v}_i$ with the expansion coefficients u_i. In doing so, (III.52) breaks down to a system of linear equations

$$\sum_j u_j (S_{ij} - M_{ij}) = R_i, \tag{III.53}$$

with the so-called stiffness matrix elements $S_{ij} = \int_\Omega d^3\boldsymbol{r}\ (\boldsymbol{\nabla} \times \boldsymbol{v}_i^*) \cdot 1/\mu (\boldsymbol{\nabla} \times \boldsymbol{v}_j)$ and mass matrix elements $M_{ij} = \int_\Omega d^3\boldsymbol{r}\ \omega^2 \boldsymbol{v}_i^* \varepsilon \boldsymbol{v}_j$ and suitable right-hand side elements R_i. From the solution of (III.53), we obtain the expansion coefficients u_i and, accordingly, the discrete solution $\boldsymbol{E}_h$ of the weak formulation of Maxwell's equations.

The search for subspaces $W_h \subset H^\perp(\mathrm{curl}, \Omega)$, which yield sparse matrices $S = \{S_{ij}\}$ and $M = \{M_{ij}\}$ and enable a stable solution of (III.53), is a major part of mathematics on the FEM. The interior domain Ω is represented by a discrete mesh with elements K_i: $\Omega = \cup_i K_i$. On each geometrical domain K, a space of functions P_K is chosen (in the case of FEM, these are polynomial functions of polynomial degree p). Additionally, linear functionals Σ_K on P_K must determine a unique basis of P_K, i.e. they are *unisolvent*. The triple (K, P_K, Σ_K) is called a finite element and the functionals Σ_K are its degrees of freedom.

The trick of the FEM is to operate not on each geometrical domain K_i separately, but to use a simple reference element $\hat{K}$. This usually has a simple shape and unit size. In 1D, for instance, $\hat{K} = (0, 1)$. Assembling of the stiffness and mass matrix of (III.53) is done on $\hat{K}$. Transformation rules map quantities on the domains K_i to $\hat{K}$. In the case of Maxwell's equations, the material mapping of the permittivity and the permeability is

$$\hat{\varepsilon} = |J| J^{-1} \varepsilon J^{-T} \tag{III.54}$$

$$(\hat{\mu})^{-1} = \frac{1}{|J|} J^T \mu^{-1} J, \tag{III.55}$$

with the Jacobian J.

The finite elements (K, P_K, Σ_K) are said to be W *conforming* if the corresponding global finite element space is a subset of W. The so-called Nédélec Finite Elements are $H^\perp(\mathrm{curl}, \Omega)$ conforming and are a common choice for the discretization of the FEM for Maxwell's equations. The numerical effort of a solution of the FEM is determined by the global degrees of freedom $N = \cup_i \Sigma_{K_i}$.

In the following we comment on solving Maxwell's equations in the exterior domain which was neglected in the previous section. As stated before the condition on the scattered field $\boldsymbol{E}_s$ to be outwards propagating is represented in *JCMsuite* with so-called Perfectly Matched Layers (PML). That is why we motivate their basic idea in the following: in 1D, the outgoing scattered field shows an oscillating dependence $\exp(ikx)$, where k is the wavenumber. Now, we allow the complex continuation of the real spatial variable x to a path $x(\tau) = L + (1 + i\sigma)\tau$ in the complex plane. Here, L is the one-dimensional size of the interior domain Ω, $\sigma > 0$ is a fixed numerical parameter and $\tau > 0$ is a real path parameter. Accordingly, the radiation condition (outwards propagating scattered field) can be formulated as

$$\boldsymbol{E}(x(\tau)) \to 0 \text{ for } \tau \to \infty, \tag{III.56}$$

where $\boldsymbol{E}$ is the total solution. Following this idea, we solve the exterior problem for complex continued quantities. Since this complex extension of all quantities can be traced back to the material mappings (III.54) and (III.55) [40], the weak formulation of the exterior problem reads:

Find $\boldsymbol{E}_s \in H^\perp(\mathrm{curl}, \Omega)$ such that

$$a_{\mathrm{ext}}(\boldsymbol{v}, \boldsymbol{E}_s) - \int_{\partial\Omega_{\mathrm{ext}}} d\boldsymbol{s}\ \boldsymbol{v}^* \times \left(\frac{1}{|J|} J^T \mu^{-1} J \boldsymbol{\nabla} \times \boldsymbol{E}_s \right) = 0\ \forall \boldsymbol{v} \in H^\perp(\mathrm{curl}, \Omega), \tag{III.57}$$

where

$$a_{\mathrm{ext}}(\boldsymbol{v}, \boldsymbol{E}_s) := \int_{\Omega_{\mathrm{ext}}} d^3\boldsymbol{r} \left[(\boldsymbol{\nabla} \times \boldsymbol{v}^*) \cdot \frac{1}{|J|} J^T \mu^{-1} J (\boldsymbol{\nabla} \times \boldsymbol{E}) - \omega^2 \boldsymbol{v}^* |J| J^{-1} \varepsilon J^{-T} \boldsymbol{E} \right]. \tag{III.58}$$

The discrete exterior problem is analogous to the discrete interior problem (III.52). The discretization of the exterior domain Ω_{ext} is done by a discretization of $(0, \infty)$ for τ and a suitable parameter σ. The mesh and especially the cut-off of the semi-finite interval for τ is not trivial and cannot be chosen

uniformly for all problems. Rather, it has to be adjusted to the specific behaviour of the scattered field for a certain problem.

That is why adaptive strategies are used to find a suitable PML [87]. These use *a priori* as well as *a posteriori* error indicators to bound the error introduced by the PML the tolerance of which can be controlled by the user. Note that this problem of an accurate discretization of the exterior domain does not occur in the FMM. Here, the condition for outwards propagating scattered fields is fulfilled by the fact that the solution of the homogeneous outer layers represents propagating and evanescent waves. However, extending the FMM to non-periodic geometries makes use of the concept of PML as well [30].

III.2.3 Convergence

When solving the discretized form of the weak formulation of Maxwell's equations, it should be guaranteed that the discrete solution $\boldsymbol{E}_h$ converges to the analytical solution $\boldsymbol{E}$. First of all, the following theorem analyses the link between these two solutions [16].

Theorem III.2.1 (Céa's Lemma)**.** *Let the bilinear form $a(u,v)$ be W-coercive and continuous. Let u and u_h denote the exact and approximate solutions, respectively. Then:*

$$||u-u_h|| \leq \frac{A}{\alpha} \min_{v_h \in W_h} ||u-v_h||,$$

where A and α are the continuity and coercivity constants, respectively.

Here, W-coercivity means that $a(v,v) \geq \alpha ||v||^2 \ \forall v \in W$ and continuity means $|a(u,v)| \leq A||v||\ ||u||$ $\forall u,v \in W$. Céa's Lemma states that the approximation error of the Galerkin method is bounded by the best approximation error with mesh-independent constants. Accordingly, the solution of the discretized weak formulation yields (up to a scaling constant) the best results for $\boldsymbol{E}_h \in W_h$. However, it should be noted that the sesquilinear form of the weak formulation of Maxwell's equations (III.51) is not coercive. Nevertheless, this condition can be generalized to the so-called *inf-sup condition* [16] and similar results as Céa's Lemma hold.

Next, we have to ensure that $\boldsymbol{E}_h \to \boldsymbol{E}$ for $h \to 0$. We define h to be the maximal element size of the geometrical domains K_i defined in the previous section. Furthermore, we denote the polynomial degree of the functional space P_K of the finite elements by p. Similar to the previous section, N is the number of global degrees of freedom $N = \cup_i \Sigma_{K_i}$. Additionally, we use $||u||_{H^1(\Omega)} := ||u||_1 = [\int_\Omega d^3\boldsymbol{r}(|\boldsymbol{\nabla} u|^2 + |u|^2)]^{1/2}$ and $||u||_{L^2(\Omega)} := ||u||_0 = (\int_\Omega d^3\boldsymbol{r}|u|^2)^{1/2}$ (only in this section).

For the classical FEM, the polynomial degree p is fixed and the mesh size h is uniformly decreased. For uniform h-refinement the approximation error is

$$||u-u_h||_1 \leq CN^{-\min\{p,r\}}, \tag{III.59}$$

where r depends on the regularity of the solution, i.e. $||u||_{r+1}$ needs to be bounded. We see that the approximation error is bounded by the chosen polynomial degree p and the regularity. When using uniform p-refinements, one can obtain unlimited convergence rate if there is no limit on the regularity:

$$||u-u_h||_1 \leq CN^{-r}. \tag{III.60}$$

That is why we expect faster convergence for increasing p rather than decreasing h. However, the dependence on the regularity of uniform h-refinements can be eliminated by using so-called adaptive h-refinements. Here, an *a posteriori* error indicator is used to identify domains in the mesh where the error is large (see Sec. IV.2.1 for a numerical example). In this case, the error of refining h is solely bounded by the polynomial degree:

$$||u-u_h||_1 \leq CN^{-p}. \tag{III.61}$$

Instead of refining either h or p, so-called hp-adaptive strategies (cf. next section), combine adapting both the mesh size h and the order p. This yields exponential convergence in the case of unlimited regularity, since the error can be estimated as

$$||u-u_h||_1 \leq C\, \exp\left(-\alpha N\right), \tag{III.62}$$

where $\alpha > 0$. For limited regularity, however, all strategies yield algebraic rates of convergence, since the regularity determines the constant C. Additionally, it should be noted that in the case of *a priori*

knowledge of the location of singularities so-called Optimal Initial Meshes [16] be generated. These yield exponential convergence rates in the preasymptotic range even for uniform p refinement. Hence, the FEM allows to include physical expectations in the discretization process in order to obtain fast convergence of this numerical method.

III.2.4 *hp*-Adaptivity

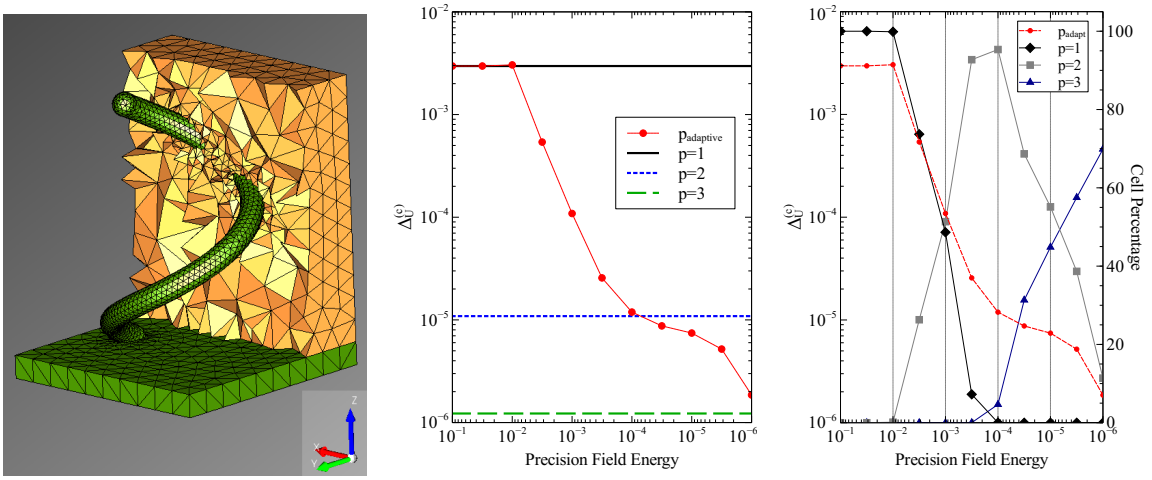
Within the scope of this work we set the polynomial degree p globally for all patches. For 3D problems this might be an inadequate choice, since one invests too much numerical effort for regions where the solution shows high regularity. On the other hand, the error can be bounded by singularities which need higher spatial resolution (cf. previous section). One possibility to deal with these local singularities is to use an *a posteriori* adaptive grid, i.e. h-refinement. An example is illustrated in Section IV.2.1. However, for adaptive h-refinement a solution has to be computed first, leading to numerical overhead. That is why in the following we analyse the *a priori* p-adaptivity.

In general, the so-called *hp*-adaptivity [16] allows both: setting the polynomial degree p on each patch separately as well as refining h on the grid locally. Local error indicators are used which essentially represent the error of a plane wave propagating in the patch with the specific local material data. In doing so, one can invest higher numerical effort in domains where it is required to do so. This approach is suited particularly for complicated devices obtaining strict mesh constraints which lead to strongly fluctuating local mesh sizes (e.g. for chiral geometries, see Sec. II.3.1).

As a small example we show a structure similar to the one presented in [21]. The tetrahedral mesh used for this computation shows large differences in patch volume [Fig. III.2(a)]. For a short convergence study the device is illuminated with CPL of wavelength $\lambda \approx 3$ μm. We apply periodic boundary conditions in the x- and the y-direction and isolating PMLs in both z-directions. The unit cell has a footprint of 2×2 μm and the helix is 1 μm in height. Both the substrate and the helix have a refractive index $n = 1.5$ (green domain). Surrounding material is vacuum ($n = 1.0$, yellow domain).

Convergence for a globally defined polynomial degree p is guaranteed [Fig. III.2(b)]. Here, the target numerical result is the one for $p = 4$. Within the software package *JCMsuite* [32] the error indicator mentioned above is realized in the user interface with the so-called *PrecisionFieldEnergy*. This reflects the accuracy of computing the electric field energy of a propagating plane wave on each patch locally. For p-adaptivity on a fixed mesh we observe convergence of the field energy as well [Fig. III.2(b)]. In Figure III.2(c), the percentage of cells with a specific polynomial degree is plotted on a second axis for each simulation. The same x- and left y-axis is used as in Figure III.2(b).

For $\Delta_U^{(c)} < 10^{-2}$ more and more cells are computed with a polynomial degree of $p = 2$ [Fig. III.2(c)]. A second regime starting at $\Delta_U^{(c)} \approx 10^{-4}$ shows that $p = 3$ is needed to obtain a more accurate result. The maximal polynomial degree for all simulations is $p = 4$ and it is only needed for the most accurate results. Note that for this particular example error estimation of the *PrecisionFieldEnergy* is roughly one magnitude too conservative compared with $\Delta_U^{(c)}$ computed with the reference solution for global $p = 4$.



(a) Geometry (tetrahedrical mesh) (b) Global vs. locally adapted p (c) Local p percentage of patches

Figure III.2: *Dielectric helix comparable to the metallic helix analysed in [21]. Due to the complex device features the mesh composed of tetrahedrons shows high fluctuations in patch volume (a). In order to reduce local numerical effort, varying polynomial degrees p for each patch are used and show convergence (b). Convergence of a globally defined finite element degree are represented by horizontal lines (solid black for $p = 1$, dotted blue for $p = 2$ and dashed green for $p = 3$). Local p-adaptivity shows convergence with respect to the a priori estimated PrecisionFieldEnergy as well (red circles). Numerical effort is determined by the percentage of cells of high polynomial degree. Their number grows for higher requested accuracy (c). For an error of the electric field energy of less than 10^{-2}, $p = 1$ (black diamonds) is sufficient. However, for more accurate results more cells need $p = 2$ (grey squares). For errors smaller than 10^{-4}, more finite elements with $p = 3$ (blue triangles) are required. Lines are a guide to the eye.*

III.3 Layering Algorithm

Standard FMM is formulated for stratified media, i.e. the geometrical description consists of layers in the z-direction with piecewise constant permittivities in the xy-plane. In contrast, FEM uses general discretizations of the geometry which are formed by patches such as tetrahedrons, prisms, bricks and pyramids. In order to work on the same geometrical description, a layering algorithm is implemented to obtain layers from general FEM grids. This is included as a post process in the software package *JCMsuite* [32] which is used for FEM simulations of the convergence study at hand. The basic idea will be outlined in the following.

First of all, layers consist of a two-dimensional *cross section*. A cross section is obtained by cutting the discretized *grid* at the z-coordinate *zCut* parallel to the xy-plane. This is done with the function *GetCrossSection(zCut, grid)*: All *faces* (edges in 2D and planes in 3D) are checked whether they intersect with the cutting plane at zCut. Additionally, only those faces are considered which are either part of the periodic boundaries or over which the permittivity (or the *domain*) changes.

The different domains are associated with a domain *id*. If one of the conditions above is fulfilled the corresponding face will be further analysed with the help of the method *GetIntersection(zCut, face)*, yielding *intersections* (points in 2D and edges in 3D). These intersections are classified with respect to their surrounding materials, i.e. their ids. Finally, closed polygons of the shapes described in the previous section are obtained by connecting the correct intersections with the help of the function *CollectPolygons(intersections)*. This procedure is summarized in the following pseudo code.

```
GetCrossSection(zCut,grid):

intersections = []
for all faces in grid
        if face intersects zCut and (face at boundary or domain changes over face)
                intersections.insert(GetIntersection(zCut,face),id1)
                intersections.insert(GetIntersection(zCut,face),id2)
return CollectPolygons(intersections)
```

For 2D problems, intersections are points and by sorting these in ascending order the piecewise constant permittivity is obtained by simply connecting one point to its successor. However, for 3D problems, there are different types of intersections which can be formed by the cut of one face and the cutting plane. The intersections are computed and classified by *GetIntersection(zCut, face)*.

First, each bounding *edge* of the face is checked whether it intersects with the cutting plane at zCut. The cutting points are obtained by *GetCuttingPoint(zCut, edge)*. If an edge is intersected at one of the end points, the intersection is said to be a *point cut*. These are ignored since other faces of the patches contribute to the polygons in the cross section. A *normal cut* is the standard case: two bounding edges have one cutting point with the cutting plane each. These two points form a possible edge of a polygon of a shape in the cross section. If the edge lies in the cutting plane, the intersection is a so-called *singular cut*. Then the bounding edge itself is a possible edge of shape-polygon. The following pseudo code displays the course of steps.

```
GetIntersection(zCut, face):

cutPoints = []
for all edges of face
        if edge intersects zCut
                cutPoints.insert(GetCuttingPoint(zCut, edge)
check case of intersection: normal / singular / point cut
return intersection
```

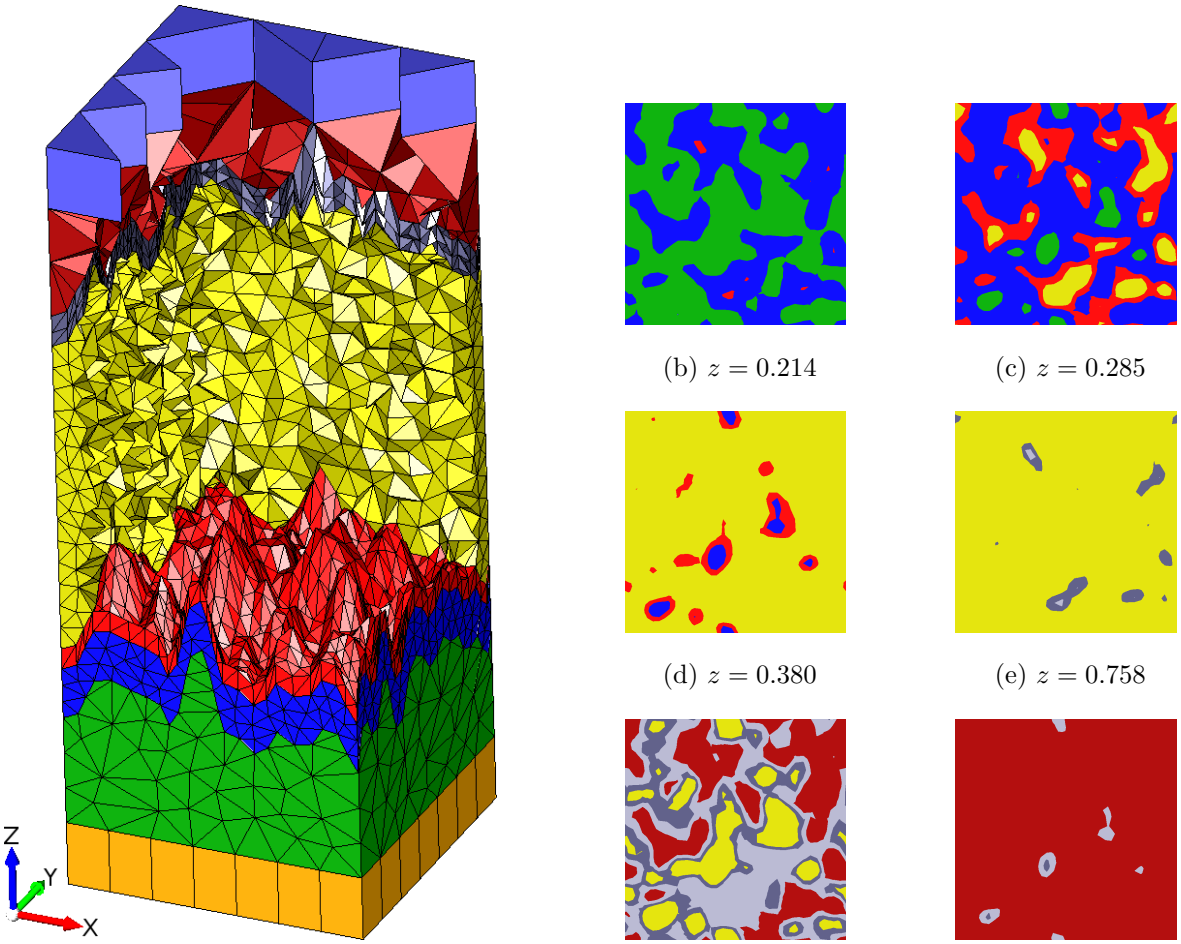
Finally, closed polygons are obtained from the set of intersections for each domain *id*: the first intersection is taken from the set of intersections. This is the first edge of a possible polygon. As long as this polygon is not closed, all intersections that are connected to the last polygon edge are possible *candidates* for closing the polygon. From this set of candidates the one is chosen which forms the smallest angle with the last edge by *GetSmallestAngle(candidates)*. In order to select the correct polygons from the sets for different domain ids, only counter-clockwise oriented ones are used. The orientation is obtained from the signed area of the polygon [90]. The intersections which form the closed polygon are deleted from the set of intersections. When all intersections are analysed, all polygons of the cross section are found. The method for this procedure is *CollectPolygons(intersections)*:

```
CollectPolygons(intersections):

crossSection = []
while !isEmpty(intersections)
        polygon = []
        polygon.insert(intersections.begin())
        while !isClosed(curPolygon)
                candidates = GetConnectedIntersections(polygon.end(), intersections)
                polygon.insert(GetSmallestAngle(candidates)
        if polygon is counter-clockwise oriented
                crossSection.insert(polygon)
return crossSection
```

The layering algorithm is one part of the unification of the interfaces of the software packages S^4 and *JCMsuite*. Its aim is to be able to systematically study the convergence of the FMM compared with the FEM. The interface is further described in Appendix A. An example of applying the layering algorithm to an arbitrary FEM grid is shown in Figure III.3. Here, the cross sections of a sequence of rough surfaces are obtained from a tetrahedral mesh.

A layer consists of the cross section, i.e. the shapes (polygons) which define the geometrical setup. Additionally, the layer has a thickness t. The thickness of the layer, which belongs to the cross section at $z = zCut$, can be computed with the option *ThicknessAdaptivity* of the layering algorithm. Here, the thickness of a layer is either defined by the distance in the z-direction between two cuts ($t = |zCut_1 - zCut_2|$) or by the extents $h_{z,i} = u_{z,i} - l_{z,i}$ in the z-direction of the material domains in which the cross section is located. The extents of material domains, sorted in ascending order in z, can be obtained with the option *DomainAdaptivity*. Then cuts are automatically set in the centre of the intervalls, which are defined by the lower $l_{z,i}$ and upper $u_{z,i}$ bounds of the extents of the material domains.



(a) *Tetrahedral mesh*

(b) $z = 0.214$

(c) $z = 0.285$

(d) $z = 0.380$

(e) $z = 0.758$

(f) $z = 0.858$

(g) $z = 0.956$

Figure III.3: *Application of the layering algorithm to a sequence of rough surfaces which are part of a solar cell setup (cf. [35] for details). The cross sections of the tetrahedral mesh (a) are obtained with the option DomainAdaptivty (cf. main text). This yields 15 automatically set cut positions. For illustration, six cross sections are displayed [(b)-(g)]. The z-coordinates are normalized with respect to the bounding box in the z-direction of the full mesh. Different colours refer to different material domains.*

Chapter IV

Simulation Results

Simulations throughout this work are carried out with the open-source solver S^4 (Stanford Stratified Structure Solver) [70] for FMM and the commercial software *JCMsuite* [32] for FEM, respectively. For the purpose of the convergence study at hand, the interface of the two solvers has been unified. This is mainly done by using consistent unit systems, illumination including phase shifts and working on the same geometrical representation. The transformation of the latter is done with the help of the layering algorithm described in Section III.3. The unification of the user interface is described in more detail in Appendix A and is implemented in a custom version of S^4 and as part of *JCMsuite*, respectively.

The convergence study through numerical experiments is outlined as follows: firstly, the interface is verified and basic FEM convergence is studied by analytical comparison of plane wave propagation in vacuum and at a simple material interface. In addition, the material approximation of FMM and specifically the inherent Gibbs phenomenon of this approximation is analysed in Section IV.2.1. Keeping in mind that the basis property of FMM basis functions is still not proven for metallic materials, we investigate an EUV (Extreme Ultraviolet Lithography) mask afterwards. Furthermore, we show the convergence behaviour of the representation of geometry in the stratified FMM in Section IV.3. Here, the numerical effects of the FFR (see Sec. III.1.2.3) are studied as well as the so-called staircase approximation. We complete the convergence study of FMM with the investigation on improved bases in 3D, which are described in Section III.1.4.

Note that the notation slightly changes compared to the theoretical Section III.1.2: now M denotes the total number of Fourier coefficients used, so $M = 2M' + 1$ with M' of the theoretical section in 2D simulations. This is due to the selection process of suitable 2D Fourier harmonics for 3D simulations (see Sec. III.1.4). The new M does not correspond to the summation indices used for the simplified arguments in Section III.1.2.

IV.1 Analytical Comparison

IV.1.1 Vacuum

In order to verify the comparison between the software packages in use and to get bounds for the expectable errors, we run a simple comparison with a propagating plane wave

$$\boldsymbol{E}(x) = \boldsymbol{E}_0 \exp\left(-i\boldsymbol{k}\boldsymbol{x}\right). \tag{IV.1}$$

We use a wavelength of $\lambda = 300.0$ nm and oblique incidence at $\varphi = 20°$, $\theta = 30°$ yielding $\boldsymbol{k} = (0.6204, 0.3582, 1.9681)^T \times 10^7$ and a normalized amplitude $\boldsymbol{E}_0 = (0.8138, 0.4698, -0.3420)^T$. The computational domain (CoDo) is chosen to be 1.5 μm in lateral (the x-) and 1.0 μm in vertical (the z-) direction. Discretized errors are computed on an equally spaced Cartesian grid with $N_x = 300$ and $N_z = 200$. As for all following problems, lateral boundary conditions are periodic.

Since (IV.1) is one of the basis functions of FMM, the errors of each quantity are solely due to numerical errors and hold true for a varying number of Fourier harmonics M [Fig. IV.1(a)]. On the other hand, FEM simulations are done with a polynomial degree $p = 3$ and show nearly exponential

convergence with the mesh size h:

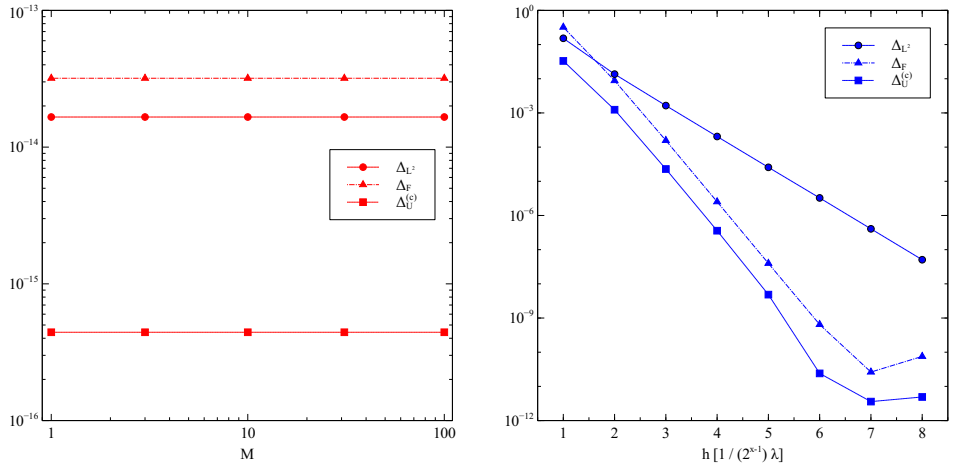
$$\Delta_k \propto \exp(-x) \tag{IV.2}$$
$$\log(\Delta_k) \approx c_k x + b_k, \tag{IV.3}$$

where $h = 1/2^{x-1}\lambda$ and the corresponding convergence rates are

$$c_{L^2} \approx -2.134 \tag{IV.4}$$
$$c_F \approx -4.034 \tag{IV.5}$$
$$c_U^{(c)} \approx -4.193. \tag{IV.6}$$

For these linear fits only the first six data points are used since the numerical error saturates for smaller side length constraints. It can be clearly seen that the near-field error Δ_{L^2} converges more slowly than the far-field and integral errors, respectively.

With the help of an automatic PML refinement scheme [87] (which is part of standard *JCMsuite*) a sufficiently discretized PML is computed and fixed for all simulations. With the help of a uniform grid refinement, the triangular grid with a side length constraint of h is refined from $h = \lambda$ to $h = 1/128\lambda$. These 2D results are displayed in Figure IV.1(b).



(a) FMM (b) FEM

Figure IV.1: *Convergence of FMM (a) and FEM (b) simulations for a propagating plane wave in 2D. Relative errors of the near-field in L^2 norm Δ_{L^2} (circles), summed relative errors of the Fourier coefficients Δ_F (triangles) and relative errors of the electric field energy $\Delta_U^{(c)}$ (squares) are shown. For their definitions refer to Section II.4. Errors are computed with respect to the analytical values. Lines are a guide to the eye. Since the plane wave is a basis function of FMM, only numerical errors contribute to the convergence with the number of Fourier harmonics M. For the FEM we observe close to exponential convergence with the grid size parameter h.*

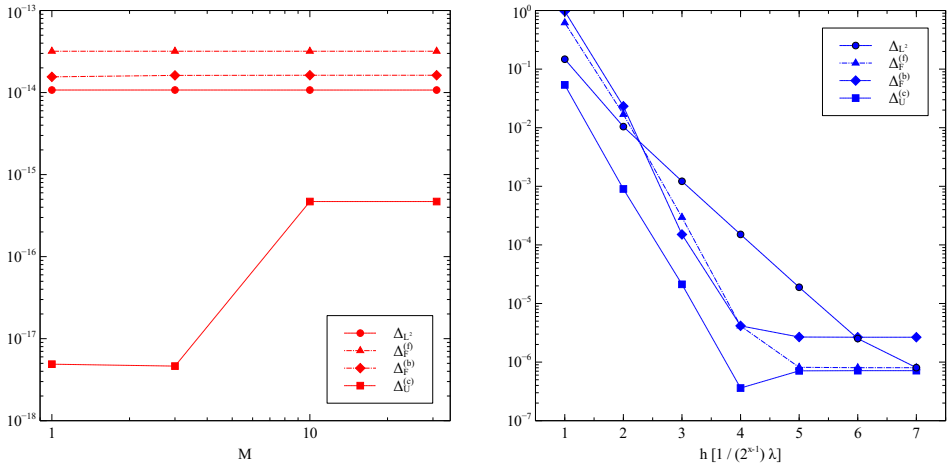
IV.1.2 Material Interface

The same setup as in the previous section is chosen to verify the numerical representation of Fresnel's equations (II.22)-(II.25) at a material interface for the given software interface. The vertical dimension of the CoDo is split into two halves of vacuum and a material with refractive index $n = 2.04$, respectively. Again, the automatic PML refinement yields a sufficient discretization in the exterior domain for FEM and the results are comparable to those of the previous section (Fig. IV.2).

The convergence rates of Eq. (IV.3) are

$$c_{L^2} \approx -2.280 \tag{IV.7}$$
$$c_F^{(f)} \approx -3.978 \tag{IV.8}$$
$$c_F^{(b)} \approx -4.212 \tag{IV.9}$$
$$c_U^{(c)} \approx -3.948, \tag{IV.10}$$

where $c_F^{(f)}$ and $c_F^{(b)}$ correspond to forward and backward propagating Fourier coeffiecients. Only the first four data points have been used here, since the numerical error saturates afterwards. Comparable with the previous section, the near-field error in L^2 norm does not converge as fast as the errors of the Fourier Transform and the electric field energy, respectively. Similar results for the propagating plane wave in 3D are obtained using the provided software interface (see Appendix A).



(a) FMM (b) FEM

Figure IV.2: *Convergence of FMM (a) and FEM (b) simulations for a propagating plane wave at a material interface in 2D. Relative errors of the near-field in L^2 norm Δ_{L^2} (circles), summed relative errors of the forward propagating Fourier coefficients $\Delta_F^{(f)}$ (triangles) as well as the reflected Fourier coefficients $\Delta_F^{(b)}$ (diamonds) and relative errors of the electric field energy $\Delta_U^{(c)}$ (squares) are shown. For their definitions refer to Section II.4. Errors are computed with respect to the analytical values. Deviations of the electric field energy ($\Delta_U^{(c)}$) of the FMM for different numbers of harmonics are pure numerical artefacts.*

IV.2 Material Approximation

IV.2.1 Fourier Series Representation

IV.2.1.1 Fast Fourier Transform and Lanczos Smoothing

To gain a first insight into the convergence behaviour of the FMM, we analyse a simple line mask with absorbing material. We use an electric field in TE polarization since it shows no discontinuities in 2D (Fig. IV.3). Geometric and illumination parameters are similar to Table 1 (data set 4) in [13]: $p_x = 800$ nm, $w = 400$ nm, $h = 65.4$ nm, $n_1 = 2.52 + 0.596i$, $n_2 = 1.56306$, $n_3 = 1.0$ (Fig. IV.3). Illumination is a plane wave with $\lambda_0 = 193$ nm propagating in upwards (the z-) direction.

For FEM results we use $p = 3$ and a first roughly discretized grid with side length constraint λ_0. With these non-optimal numerical parameters, FEM converges well (cf. Table 2 in [13]). Firstly, we use a naive global refinement of each patch as in Section IV.1 and, secondly, an adaptive refinement [28] (which is part of standard *JCMsuite*) to reduce numerical effort. Both the near-field error and the error of the 0-th order Fourier coefficient converge faster for the adaptive than for the uniform refinement strategy (Fig. IV.4). Although the order of convergence is comparable, values of the relative errors differ by more than two orders of magnitude. In particular, the far-field error is drastically reduced by

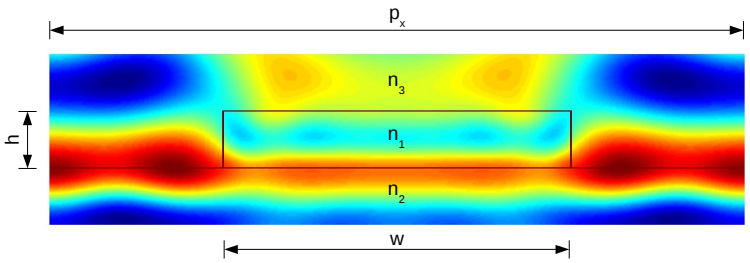


Figure IV.3: *Simple line mask with absorbing material. Geometric parameters are similar to Table 1 (data set 4) in [13]. The structure is illuminated with a perpendicularly propagating plane wave of wavelength $\lambda_0 = 193$ nm in TE polarization. Near-field intensity distribution is shown in linear colour scaling. Features of the structure are depicted by red lines.*

using an adaptive grid refinement: *h*-adaptivity leads to a much smaller number of unknowns for the same error level. Although it produces some numerical overhead it is very useful for field singularities such as plasmonic effects or sharp metallic edges [40].

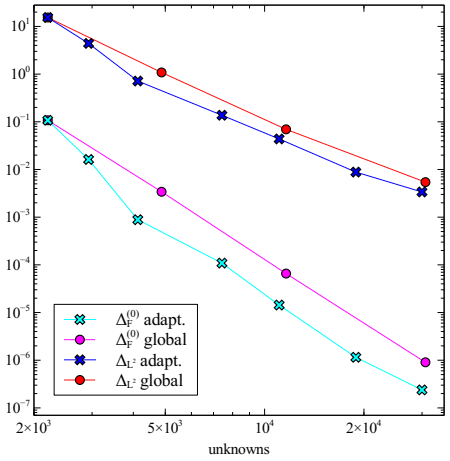


Figure IV.4: *Convergence of the FEM for a line mask with absorbing material. Global refinement steps (circles) are compared with h-adaptive refinement steps (crosses). The latter leading to reduced numerical effort for a specific error constraint. The near-field (Δ_{L^2}, red and blue) has an error roughly more than two orders of magnitude greater than the 0-th order Fourier coefficient ($\Delta_F^{(0)}$, magenta and cyan) for this structure.*

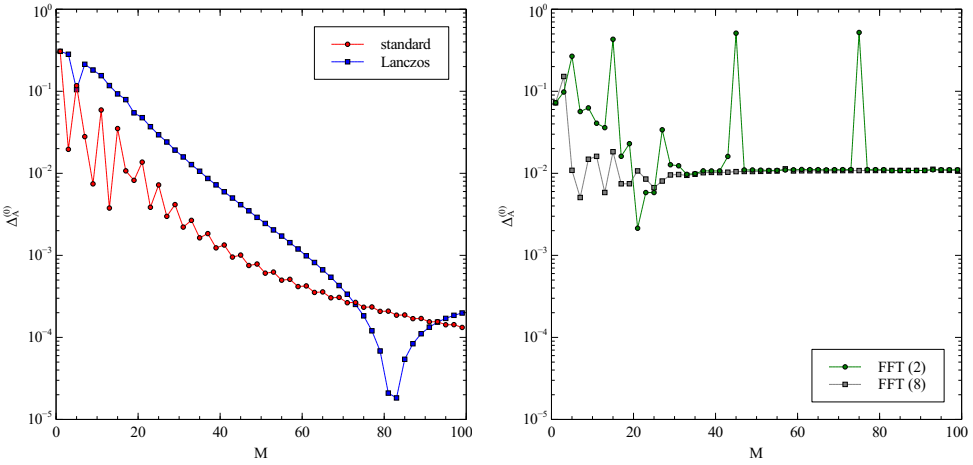
In Figure IV.5 we analyse standard FMM, a smoothing method for the discontinuous permittivity and a standard FFT approach (see Sec. III.1.4). Standard FMM, using a closed-form Fourier Transform of the permittivity field ε, leads to equal results with a relative error smaller than 10^{-4}. The inherent convergence behaviour of FMM is oscillatory [Fig. IV.5(a)]. Most authors explain this observation with the well-known Gibbs phenomenon [34, 7]. This argument is analysed in detail in Section IV.2.2.

To reduce this effect of the Fourier Transform of a discontinuous function a well established concept is smoothing (which is also known as windowing in the context of signal processing). Possible filters include Gaussian smoothing which is used for ASR within the FMM [18]. Additionally, Lanczos [66] or subpixel [20] smoothing can be applied to discontinuous materials. The latter can be used to generate smoother vector fields for improved basis sets (see Sec. IV.4).

Here, we investigate the error of the 0-th diffraction efficiency, i.e. the absolute value of its Fourier coefficient [$\Delta_A^{(0)}$, (see Sec. II.4)]. Lanczos smoothing reduces the oscillatory effect of the discontinuity as expected [Fig. IV.5(a)]. Yet it also limits the relative error itself, leading to the requirement of more harmonics to obtain the same accuracy as for standard FMM. Most of the time this is undesirable. However, it reduces the possibility of using an oscillatoric peak of the convergence as the so-called best solution. On the contrary, problems do not occur in the asymptotic regime of FEM which can be controlled more easily because of well-known mathematics on its convergence (see Sec. III.2.3). This plays a more important role for more complex examples, e.g. for the underestimation of the errors of

3D simulations (see Sec. IV.4).

Including the phase of the 0-th order Fourier coefficient ($\Delta_F^{(0)}$), Lanczos smoothing may even lead to results which are several magnitudes worse than the error of its absolute value. In conclusion, we do not observe improvement of FMM convergence using Lanczos smoothing. It even seems to corrupt correct results of standard FMM. This should be noted for future developments of the FMM mentioned in the outlooks of Chapter 4 of [49].



(a) Standard and Lanczos smoothing FMM

(b) FMM using FFT

Figure IV.5: *Convergence of the FMM for a line mask with absorbing material. Standard and Lanczos-smoothed FMM (a) as well as FFT (b) are analysed with respect to the absolute value of the 0-th order Fourier coefficient ($\Delta_A^{(0)}$). Standard FMM (red circles) shows oscillatoric convergence behaviour. Applying Lanczos smoothing (blue squares) to the discontinuous material data flattens these characteristics but leads to worse results, which is even stronger for the phase error ($\Delta_F^{(0)}$, not displayed). Using FFT requires an additional oversampling factor [numbers in brackets in the key, (b)]. Its lower bound two (green circles) leads to inaccurate results even in the saturated regime (peaks at $M = 45, 75$). A sufficient oversampling factor of eight (grey squares) reduces this effect significantly. However, an inherent error bound of approximately 10^{-2} when using the FFT for this example is observed.*

Another standard method to significantly increase speed of a numerical Fourier Transform is the so-called Fast Fourier Transform (FFT) [65]. In the context of FMM, however, it does not yield faster results (see Sec. III.1.4). To obtain comparable results to closed-form FT, an oversampling factor is introduced. To satisfy Nyquist's Sampling Theorem [65] an oversampling factor of at least two has to be chosen. We use this lower bound as well as a common sampling resolution of eight, to analyse convergence with respect to FFT [Fig. IV.5(b)]. The oversampling factor is multiplied with the largest reciprocal lattice integer used for the specific choice of M harmonics. In practice, convergence behaviour before saturation is influenced strongly: For small oversampling errors increase, while for high oversampling errors are reduced. In the former case we even observe peaks in the saturation regime [$M = 45, 75$, green circles in Figure IV.5(b)].

These peaks vanish when using higher oversampling, namely a factor of eight. In this case, both pre- and post-asymptotic behaviour is flattened. Nevertheless, for the specific example we observe saturation of the FFT results compared to the closed-form Fourier Transform at about 10^{-2} [note equal axis scaling in Figure IV.5(a) and IV.5(b)]. This does not meet expectations and should be further analysed for a broader range of examples.

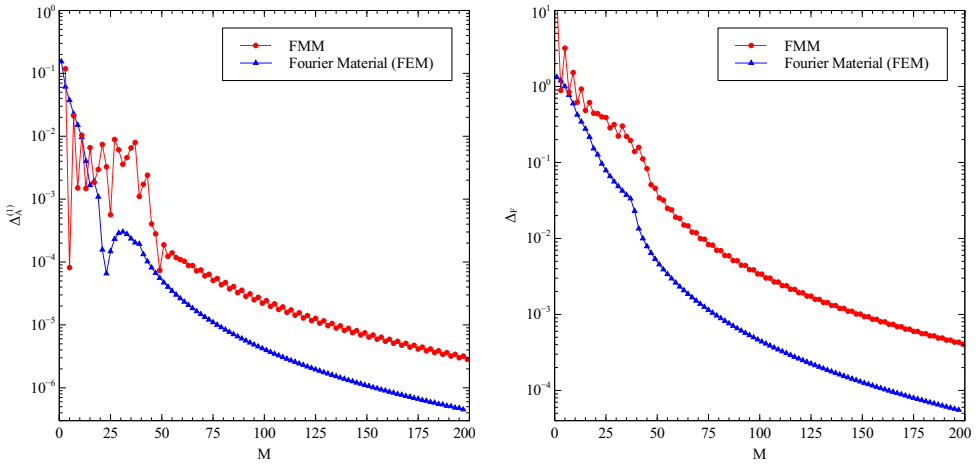
To conclude, this investigation of FFT may yield much faster results yet introduces an unexpected error bound. Interestingely, lower and upper error bounds are comparable for the phase included error $\Delta_F^{(0)}$ which is in contrast to the observation of Lanczos smoothing in the previous paragraphs.

IV.2.2 Gibbs Phenomenon

When using a plane wave basis for representing discontinuous functions, the first thing that limits the accuracy of a finite basis set is the well-known Gibbs phenomenon [29]. However, Li showed by his Inverse Rule that the convergence in TM polarization of the first formulations of the FMM was mainly

constrained by convolution in Fourier space, rather than Gibbs spatial overshoots of the permittivity profile. He states that the convergence problems of FMM are solved by this multiplication rule of Fourier series [7]. On the other hand, recent research still finds the Gibbs phenomenon to be a limiting factor of this method [34] (see Sec. IV.2.3).

In order to analyse the influence of the Fourier series representation of the index profile in 2D, we use a dielectric binary grating similar to the one in [56]. Illuminating the structure in TE polarization separates the problem of Fourier convolution from the Fourier series representation itself. We use a vacuum wavelength of $\lambda_0 = 300.0$ nm and an incident polar angle $\theta = 10°$. The pitch is $\Lambda = 10\lambda_0$, grating height $d = 0.5\lambda_0$ and the grating width $w = 0.5\Lambda$. The refractive index of grating and substrate is $n = 2.04$ and illumination is from free space with $n_f = 1.0$ (cf. Fig. IV.3: $\Lambda = p_x$, $d = h$, $w = w$, $n = n_1 = n_2$, $n_f = n_3$. Note mirroring at x-axis, since illumination is from free space). Investigations of the Inverse Rule, i.e. comparing TE and TM illumination is done in Section IV.3 and Figure IV.12, respectively.



(a) 1st diffraction order

(b) all Fourier coefficients

Figure IV.6: *Analysis of the influence of the Fourier Transform of the permittivity profile. Simulations of the conventional FMM (red circles) are compared with those obtained by using the analytic Fourier series of the permittivity in the grating layer in FEM computations with four h-adaptive refinement steps (blue triangles). Reference solution is a fully converged FEM simulation for the discontinuous dielectric binary grating introduced in [56]. Considering energy transported in the first diffraction order, FMM yields fast and accurate results (a). In contrast, the ability to represent a correct far-field pattern is limited by the total error of all complex Fourier coefficients Δ_F (b). Here, for small as well as for high numbers of harmonics full FMM leads to inaccurate results (note the different y-axis scaling). In both cases FMM leads asymptotically to roughly one order of magnitude worse results than those which could be obtained with a Fourier series representation of only the material (and not the electromagnetic fields as well).*

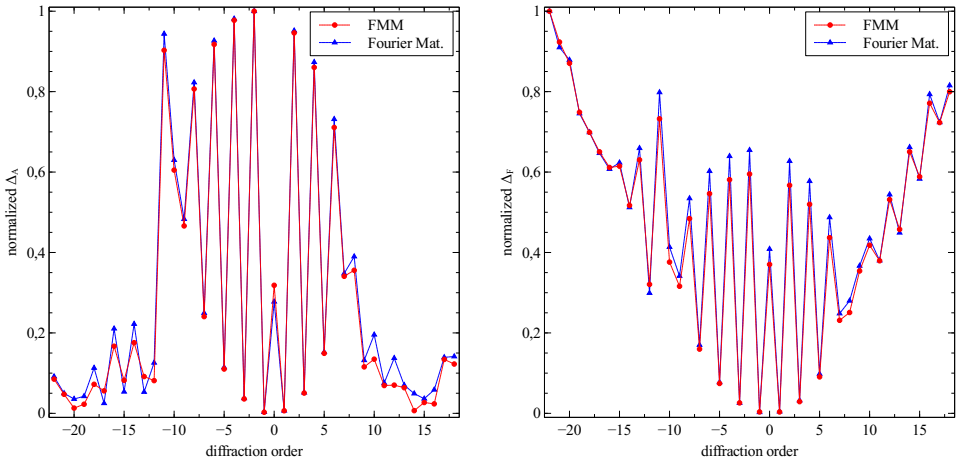
First, we show convergence behaviour of the first diffraction order in more detail than in Figure 3 of [56]. Convergence of the absolute value of the first order Fourier coefficient is displayed in Figure IV.6(a). Errors less than 10^{-5} can be obtained with only 200 harmonics in the basis set. The reference solution is again a well converged FEM result.

Separation of the material approximation with plane waves from the plane wave basis for the electromagnetic fields is here not achieved by an oversampling in the field basis as in [34]. Rather, we use FEM with its proven convergence characteristics, to represent analytically the permittivity profile in a Fourier basis in the grating layer. We use an initial grid with a discretization of half the wavelength of the shortest Fourier basis function in the grating layer and half the material dependent wavelength in the homogeneous layers. Additionally, we use the h-adaptive refinement strategy of *JCMsuite* with four refinement steps to ensure locally converged field distributions over the whole CoDo. Simulations are done with a finite element degree of $p = 3$.

The results using this Fourier representation of the permittivity profile on one side and using a polynomial basis for the fields on the other side are also shown in Figure IV.6(a). They suggest that pre-asymptotically the FMM yields better results for the energy propagated by the first diffraction order. Yet asymptotic behaviour shows a more inaccurate result when using the plane wave basis for the fields of approximately one order of magnitude. This counter-intuitively good convergence

of propagated energy often occurs in analysing the FMM (see Sec. IV.3) and many authors use it to demonstrate that FMM yields accurate results. However, they miss the inaccurate phase of the Fourier coefficients when using FMM.

By including the phase correlations of the complex Fourier coefficients [Fig. IV.6(b)] this picture is clarified: phases of the Fourier coefficients are not correctly computed using the FMM and do not even reach results which would be obtained using a Fourier Transform of only the grating layer material. Furthermore, we find that the typical sinusoidal convergence behaviour of FMM is not due to the Gibbs overshoots but rather due to the plane wave basis itself. This behaviour is often shown on linear plots of the figure of merit throughout FMM literature and this scaling does not allow further insight into the detailed accuracy of the method.



(a) absolute value of Fourier coefficients

(b) complex Fourier coefficients

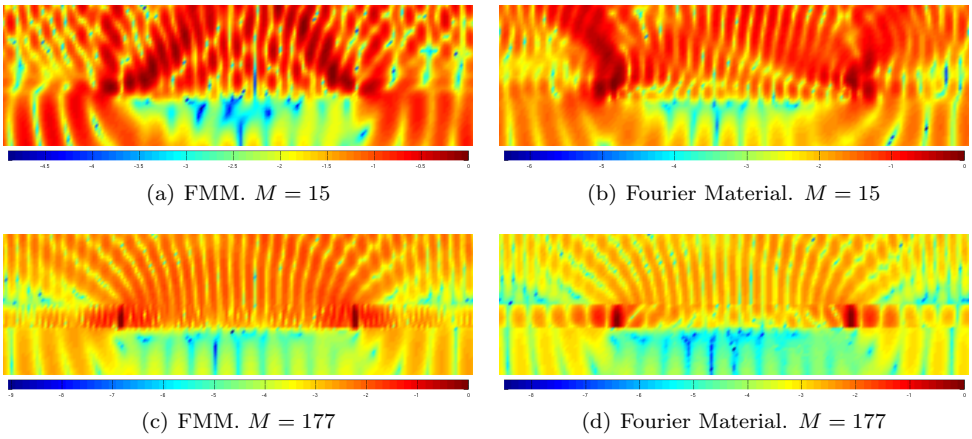
Figure IV.7: *Errors of the diffraction orders for $M = 197$ of the full FMM (red circles) vs. contributions solely due to the Fourier Transform of the permittivity in the grating layer (blue triangles). Errors in the ± 1st diffraction order are minimas leading to optimistic conclusions of the convergence of FMM in [56]. Additionally, Eq. (II.61) is confirmed for this example: Errors of high diffraction orders of the complex Fourier coefficients (b) increase from the 0-th diffraction order, but decrease for their absolute values (a). In conclusion, FMM is suitable for energy (conservation) analysis but does not represent the full far-field pattern including phase correlations accurately.*

In physics simulated with the FMM one is often only interested in energy diffraction [77]. In addition, in diffraction theory only diffraction efficiencies are investigated. For these purposes FMM yields accurate enough results because of Eq. (II.61). Figure IV.7 shows this statement in more detail for the real world problem of this specific binary grating. We see that, first of all, the ± 1st diffraction orders show minimal errors for e.g. $M = 197$ Fourier harmonics [Fig. IV.7(a)]. This holds since the grating is designed to optimize diffraction into the first order. We see that both the full FMM and the Fourier representation of the material show decreasing relative errors for higher diffraction orders, i.e. results concerning energy transportation are accurate.

In contrast, the relative error of the complex Fourier coefficients increases for higher diffraction orders [Fig. IV.7(b)]. This severely limits applicability of the FMM for problems for which accurate far-field patterns need to be resolved, e.g. in metrology. Although absolute relative errors are higher for the full FMM as stated in the previous paragraphs, surprisingly the normalized error behaviour is in many diffraction orders better than the total Fourier representation.

Finishing the analysis of the material approximation with a simple Fourier Transform of the relative permittivity in 2D, we take a closer look at near-field error contributions. This is done with the help of the previously defined local relative L^2 error $\Delta_{L^2}(x)$ (Def. II.4.4). This normalized error shows areas where near-fields are not well approximated and possible errors of the far-field reconstruction from the Fourier coefficients arise from.

We compare local error characeristics on a $\log_{10}$ scaling (Fig. IV.8). First of all, pattern features of the full FMM and the Fourier representation of the grating layer are comparable as expected. Evolution of errors with the number of harmonics are shown for two examples at $M = 15$ and $M = 177$. The pitch and the illumination angle of this device limit the number of propagating diffraction orders to 41. Using fewer plane waves for the FMM, namely $M = 15$, leads to relative errors less than 10^{-2} for the figure of merit $\Delta_A^{(1)}$ [Fig. IV.6(a)].



(a) FMM. $M = 15$ (b) Fourier Material. $M = 15$

(c) FMM. $M = 177$ (d) Fourier Material. $M = 177$

Figure IV.8: *Near-field error contributions in* $\log_{10}$ *scaling of the local relative* L^2 *error* $\Delta_{L^2}(x)$*. Local error origins of the full FMM (a, c) are compared to those of the Fourier Transform of the grating layer material functions in FEM simulations (b, d) for two different numbers of plane wave basis functions,* $M = 15$ *and* $M = 177$*, respectively. FMM's Fourier basis introduces more oscillatoric error patterns than the material itself. Specifically, not only lateral sinusoidal errors are obtained but also vertical oscillations occur in the upper homogeneous layer. In addition, higher oscillations in the grating layer are observed (c). Errors arising in lower layers are propagated through the structure via the coupling of the scattering matrix algorithm. This leads to a smoother error distribution from one dielectric to another dielectric layer which does not show error discontinuities at layer interfaces introduced by the Fourier material (d) (cf. different behaviour when metallic features are involved: Fig. IV.11).*

On the other hand, analysing the locally defined L^2 error reveals high error contributions in the homogeneous substrate layer in the propagation direction $+z$ [Fig. IV.8(a)]. These features are much stronger for the full FMM than the error caused solely by the permittivity's Fourier series [Fig. IV.8(b)]. Additionally, in the upper layer more complicated patterns are introduced by the Fourier basis: not only lateral sinusoidal behaviour is observed but an additional vertical oscillatoric characteristic can be seen. Errors in the near-field approximation are mostly not limited to the locations of the permittivity discontinuities as for the Fourier material. The errors of the full FMM propagate in triangular shape through the substrate layer and decrease with a high depth of penetration. This shows the difficulty of error propagation through insufficiently resolved layers across the whole CoDo of the FMM by the coupling through the scattering matrix algorithm (cf. metallic layers in Section IV.2.3).

For a much higher number of basis plane waves than propagating diffraction orders, the overall pattern of the full FMM is comparable to the error caused by the Fourier Transform of the index profile [Fig. IV.8(c)]. In particular, the number of error fringes introduced in the upper homogeneous layer is the same [Fig. IV.8(d)]. So discontinuities due to the Fourier material at the layer interfaces are not correctly represented in the FMM. Again, a smoothing yet faulty coupling of the insufficiently converged layer eigenmodes is observed here.

In addition, extra oscillations of the local error are introduced in the grating layer. These do not arise from the permittivity profile, which shows smoother errors, but from the plane wave basis. This could potentially be lowered by using a layer specific number of harmonics to consider more complex layer structures compared with, for example, homogeneous ones (cf. proposal in next section). It should be noted that errors at the discontinuities themselves are more localized for the FMM which could be caused by the additional oscillations of the plane wave basis leading to smaller spatial error wavelengths.

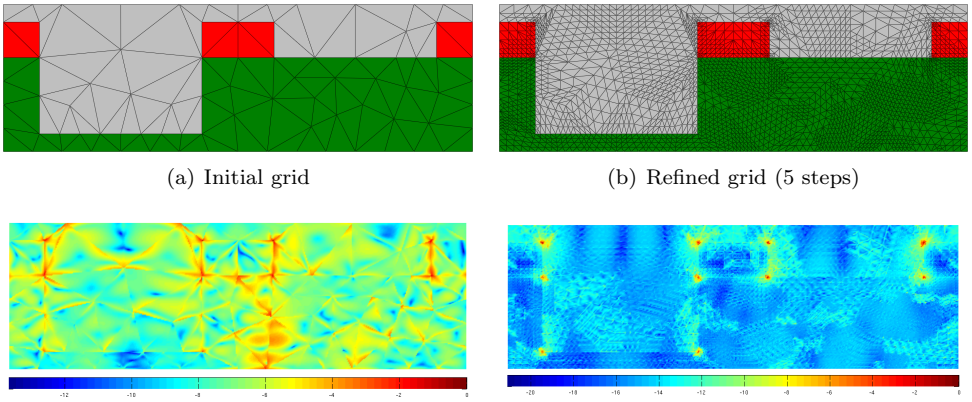
Overall, we have shown in the analysis of a TE illuminated dielectric binary grating that the Fourier representation of the material functions, i.e. the Gibbs phenomenon, is not the only contribution to FMM errors. Additional oscillatoric error input can be attributed to the plane wave basis. We showed that errors arising in the first layers propagate through the whole device. This should be noted with respect to the staircasing approximation which introduces a high number of additional layers (see Sec. IV.3.2). Furthermore, we showed that conventional energy investigations, i.e. studies of the absolute value of Fourier coefficients, are sufficiently and quickly approximated by the FMM. For correct far-field patterns, however, the phases of these coefficients need to be computed with a much higher number of harmonics, especially for high diffraction orders. This requirement for a higher number of harmonics should be taken into account for convergence studies such as [48].

IV.2.3 Metallic Scatterers

Analysing the properties of nano-optical devices which include metallic features is of great interest. Although in the beginning of simulations with the FMM (see Sec. III.1.1) mostly dielectric gratings were computed, the fields of lithography and especially plasmonics deal with complex refractive index profiles. Surface Plasmons only occur when an electromagnetic field interacts with metals [59]. This near-field effect is expected to be controllable with the help of, for example, chiral particles (see Sec. II.3.1). Additionally, within the fabrication process of integrated circuits and chips metallic masks are used. An example of such an EUV mask is analysed in this section as a representative structure for the FMM with metallic materials.

The eigenmode basis of FMM is proven to form a complete set for real index profiles, i.e. dielectrics. Yet a formal proof of its completeness in the context of complex index profiles is still lacking and usually the plane wave basis is simply assumed to be suitable for these structures as well [39].

Convergence problems of FMM for metallic devices are reported and analysed by Kim et. al [34], among others. They modify FMM to use more plane waves in the basis set than for the representation of the permittivity profile. In the conventional FMM the number of harmonics of the basis is strictly coupled to the number used for the Fourier Transform of ε. These authors conclude that the convergence problems originate from the Gibbs phenomenon and from non-convergence of highly evanescent eigenmodes. This conclusion (yet for dielectrics) was analysed in the previous section and at least the oscillatoric fluctuations are shown to not originate from the Gibbs phenomenon. Here, we make a convergence study for metallic features and show that near-field quantities are in fact strongly limited for these kinds of absorbing materials. However, convergence for the far-field Fourier Transform is seen to behave much better in the given limits.



(a) Initial grid (b) Refined grid (5 steps)

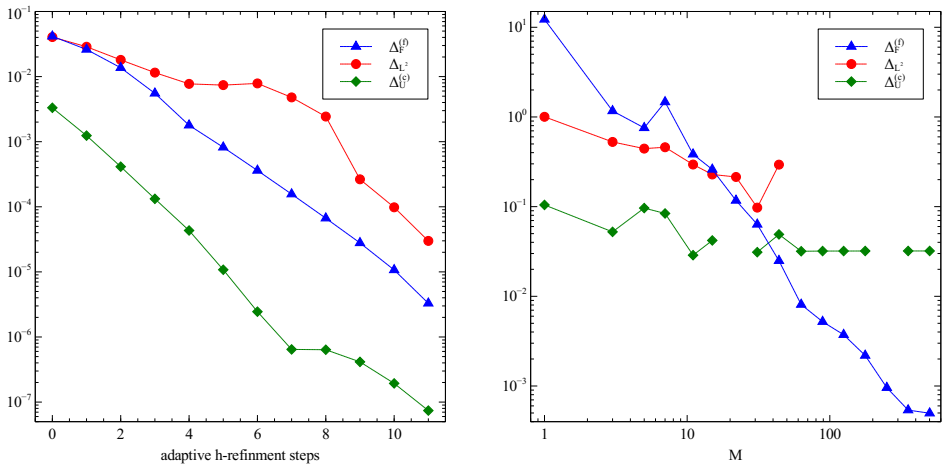
(c) Initial local $\Delta_{L^2}(x)$ (d) Refined local $\Delta_{L^2}(x)$

Figure IV.9: *Adaptive grid refinement using FEM for an EUV mask including metallic scatterers. The initial grid has side length constraints of roughly $\lambda/2$ (a) and is iteratively locally refined using JCMsuite's h-adaptivity strategy [the refined grid after five steps is shown in (b)]. The local normalized L^2 error $\Delta_{L^2}(x)$ is plotted in* ln *scaling [(c) and (d)]. As expected, this refinement localizes errors at the metallic corners where field enhancements occur.*

The best case illumination for FMM, i.e. TE polarization, is used with the standard EUV wavelength of $\lambda_0 = 193.0$ nm. Figure IV.9(a) depicts the mask. It is composed of a substrate with refractive index $n = 1.563$ (lower green domain), two metallic scatterers with $n = 0.842 + 1.647i$ (red domains) and the superstrate which is composed of air ($n = 1.0$, upper grey domain). The groove in the substrate is 172 nm deep and the scatterers are 80 nm high and 160 nm wide. Periodic pitch of the mask is 1.04 μm.

Note that it is expected that FMM does not converge accurately in TM polarization for this structure: The Inverse Rule is only applicable for certain permittivity profiles (cf. conditions on $1/f$ in Theorem III.1.4). None of these conditions is fulfilled for the layers including the metallic scatterers. So here only TE polarization can be simulated efficiently with the FMM.

For FEM calculations we use an *a posteriori h*-adaptive grid refinement as in Section IV.2.1.1. This is particularly suitable for the expected field enhancements at the corners of the metallic scatterers. This approach narrows the local L^2 error $\Delta_{L^2}(x)$ down to these spots of field singularities (Fig. IV.9). In doing so, we obtain well converged results for the forward propagating Fourier coefficients ($\Delta_F^{(f)}$), the local field (Δ_{L^2}) as well as the energy stored in the CoDo ($\Delta_U^{(c)}$ [Fig. IV.10(a)].



(a) FEM (b) FMM

Figure IV.10: *Convergence characeristics for metallic EUV mask. The h-adaptivity (cf. Figure IV.9) leads to well converged FEM results (a) for both the near-field [Δ_{L^2} (red circles), $\Delta_U^{(c)}$ (green diamonds)] and far-field [$\Delta_F^{(f)}$ (blue triangles)]. FMM errors are several magnitudes greater (b) and especially near-field quantities are of limited accuracy. Nevertheless, forward propagating Fourier coefficients still converge for a medium number of harmonics M. Missing values for L^2 and energy errors result from either numerical instabilities or a disproportional numerical effort evaluating near-fields of the FMM. This could be further optimized in the FMM implementation.*

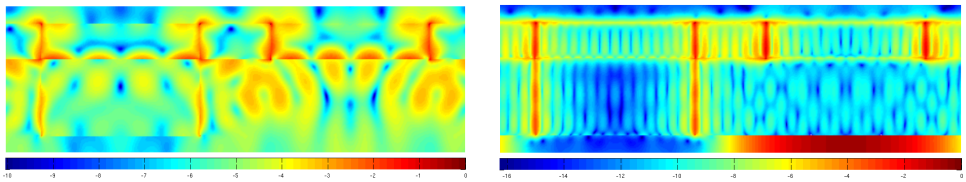
On the other hand, we see that near-field FMM results, namely energy and field distribution, are limited, while the Fourier coefficients converge slowly but monotonously [Fig. IV.10(b)]. Here, the error of the complex Fourier coefficients, i.e. including the phase of the far-field pattern, is investigated since it is the phase-corrected far-field which acts on photoresists in the lithography process. To obtain the correct phase correlations using FMM a high number of harmonics M is needed. For small numbers of M the FMM leads to incorrect results. This effect is particularly strong when M is smaller than the number of diffraction orders, which is determined by the pitch. For the energy propagated by the diffraction orders (Δ_A) the errors are again smaller.

Analysing the origin of the near-field errors, we consider the locally defined L^2 error $\Delta_{L^2}(x)$ (Fig. IV.11). We confirm the findings of Kim et. al [34] that the error is localized at the permittivity discontinuities and enhanced at metallic domain interfaces. Furthermore, the inherent sinusoidal pattern of the Fourier basis is clearly distinguishable [Fig. IV.11(b)]. For a medium number of harmonics, FMM results additionally show problems for the coupling from dielectric layers to those including metallic features [upper part of Figure IV.11(b)]. As expected, errors are much higher in the absorbing material.

In the lower right-hand part of Figure IV.11(b) we observe faulty back coupling behaviour even in the dielectric layers. These localized error enhancements occur for certain numbers of harmonics and are expected to originate from the non-convergent, highly evanescent modes mentioned by Kim et al.. This erroneous coupling characteristic, caused by an insufficient eigenmode basis, leads to discontinuities in the local near-field error [Fig. IV.11(a)]. Clearly, this is the origin of inaccurate results of this method. In addition, problems in the layer including the metallic scatterers are seen in this investigation.

From these findings we suggest an improvement to the FMM by using an adaptive number of harmonics in each layer. Through this, the more difficult eigenproblems of layers including e.g. metallic features could be solved using an increased number of basis functions. This would probably yield better converged results within these layers and consequently for the whole device.

In conclusion, the plane wave basis of the FMM strongly limits its application for systems including metallic materials. Although far-field results converge well up to problem-specific limits, near-field distributions show much worse characteristics which restrict interpretation of e.g. plasmonic effects obtained with this numerical method. Nevertheless, improvements can be gained by decoupling the material Fourier representation and the Fourier-like basis which is shown by Kim et. al.. Additionally, we propose using an adaptive number of harmonics in each layer separately, which could lower errors in layers with complex index profiles.



(a) $M = 11$ (b) $M = 44$

Figure IV.11: *Local near-field error contributions of the FMM for an EUV mask including metallic scatterers. An insufficient eigenmode basis (a) leads to both error contributions distributed over the whole CoDo and discontinuities in the local L^2 error $\Delta_{L^2}(x)$ which is plotted in normalized* ln *scaling. For an increasing number of harmonics, the error localizes at the permittivity discontinuities as expected (b). Error contributions at the metallic domain boundaries and inside these domains are higher than those in dielectric layers. Not only in the domains themselves are near-field errors enhanced but also at the coupling interfaces with layers including complex permittivities the near-field is not well approximated [upper part of (b)].*

IV.3 Geometry Approximation

IV.3.1 Fourier Factorization

As stated before (see Sec. III.1.2) the so-called Inverse Rule represents a major breakthrough in the convergence characteristics of the FMM. We illustrate its improvement with the help of the standard example of a dielectric binary grating analysed before (see Sec. IV.2.2).

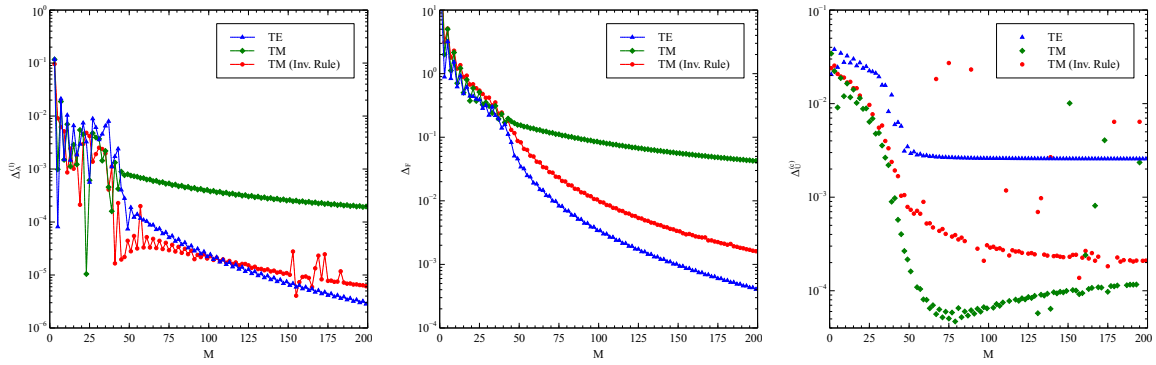
As before we use a fully converged FEM reference solution obtained with a finely discretized adapted PML and finite element degree $p = 3$. Initial grid discretization is done with a material adapted side length constraint of $1/2\lambda$ where λ is the refractive index n dependent wavelength $\lambda = \lambda_0/n$. Up to five uniform refinement steps are studied and yield a clear h-convergence with several millions of unknowns.

Our simulations confirm the well-known findings of Moharam et al. [56] that the power transported in the first diffraction order of this grating converges well for TE illumination and shows relative errors smaller than 10^{-5} for less than 200 harmonics used in the FMM computations [Fig. IV.12(a)]. Preasymptotic oscillatory behaviour is seen as well as the much slower and more inaccurate convergence for TM polarization when the standard Laurent's Rule (Def. III.1.3) is used. Using the Inverse Rule (Def. III.1.5) leads to much better results. However, the convergence rate is roughly the same and for less than 100 harmonics M, results show heavy fluctuations. The study of the absolute value of the coefficient of the first diffraction order even suggests that for a certain regime (approx. $40 < M < 100$) computations with the Inverse Rule for TM illumination are more accurate than those computed for TE polarization.

This counter-intuitive behaviour is clarified when including phase relations in the error analysis. The total error of all Fourier coefficients shows that FMM gives better results for TE polarization than expected [Fig. IV.12(b)]. Li's results of a better approximation using the Inverse Rule for the more complex problem of TM illumination still hold true for the error including phase correlations. However, the Inverse Rule does not yield better results than TE for any number of harmonics in contrast to observations made on the absolute value of the first diffraction order. Convergence characteristics of the Inverse Rule for TM are similar to those of the Laurent's Rule for TE but lead generally to more inaccurate results [note the different y-scaling in Fig. IV.12(a) and IV.12(b)].

An indicator for near-field results is the error of the energy stored in the CoDo. Its relative error shows much different results than the far-field discussion before. First, the error of energy for TE illumination saturates quite early at around $M = 50$ [Fig. IV.12(c)]. Surprisingly, TM simulations show better results in general. Although the Inverse Rule flattens oscillatoric convergence characteristics, it leads to more inaccurate results for all basis sets.

The difference of TE and TM errors could be explained by a scaling error in the implementation of either the software interface or S^4. However, this scaling error would be equal for TM simulations with and without the Inverse Rule. That is why the general assumption that the Inverse Rule yields better results also for near-field patterns needs to be questioned. Further investigation should compare the correlations of the Fourier Transform of the near-field computed with the different eigenmode basis set of the Laurent's and the Inverse Rule, respectively.



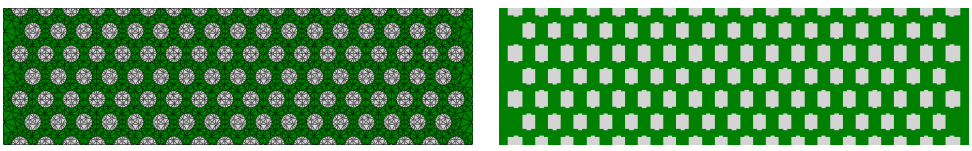
(a) 1st diffraction order (b) all Fourier coefficients (c) energy stored in CoDo

Figure IV.12: *Comparison of TE, Laurent's Rule TM and Inverse Rule TM convergence of the FMM for a dielectric binary grating (see Sec. IV.2.2). Well-known results for the energy transported in the first diffraction order (a) are confirmed: TM computations using Laurent's Rule (green diamonds) converge very slowly. This is corrected with the help of the Inverse Rule (red circles). The latter shows similar convergence characteristics as TE simulations (blue triangles). Inverse Rule TM calculations seem to yield even more accurate results for a certain number of harmonics. Including phase relations of all Fourier coefficients (b), however, shows that this is not the case. Nevertheless, it confirms the necessity of using the Inverse Rule to obtain better results for less numerical effort. Contradictory, analysing the error of the energy (c) leads to the conclusion that Laurent's Rule seems to yield more accurate near field distributions. Several magnitude difference between TE and TM results in general could result from a numerical error (cf. main text). This does not disprove the finding that the energy for Laurent's Rule TM simulations converge much faster than those with the Inverse Rule.*

IV.3.2 Staircasing

The FMM in its original formulation [56] is limited to piecewise constant permittivity profiles. This is reasonable from the viewpoint of lamellar gratings for which this method was developed. However, already the authors of this first formulation extended their method to arbitrary grating profiles using the so-called staircasing [55]. For simplified interfaces such as sinusoidal gratings specialized methods exist in grating theory, e.g. the C-method [14]. Although these should be preferred when simulating non-lamellar structures [49], in order to generalize the FMM, staircasing is a regularly used tool. In [15], staircasing is justified because structures are analysed in resonant regimes. In general, Popov et al. showed that sharp maxima are introduced into the near-field distribution at the staircasing edges [64]. In this section we analyse the staircase approximation for a Photonic Crystal (PhC) slab in 2D.

The PhC slab is adapted from a study of PhC waveguides proposed for optical sensing [4]. For these purposes vacuum rods with radius $r = 120$ nm are etched inside a dielectric background material with permittivity $\varepsilon = 12$. The unit cell is formed by a stack of 17-18 rods in the propagation direction [i.e. to the right of Figure (Fig. IV.13)]. Note the different illumination direction compared to the other structures analysed in the scope of this work. For waveguide purposes, one row of rods is removed and FMM is used with a super cell, although FMM is extended to the so-called aperiodic FMM with PMLs [30].



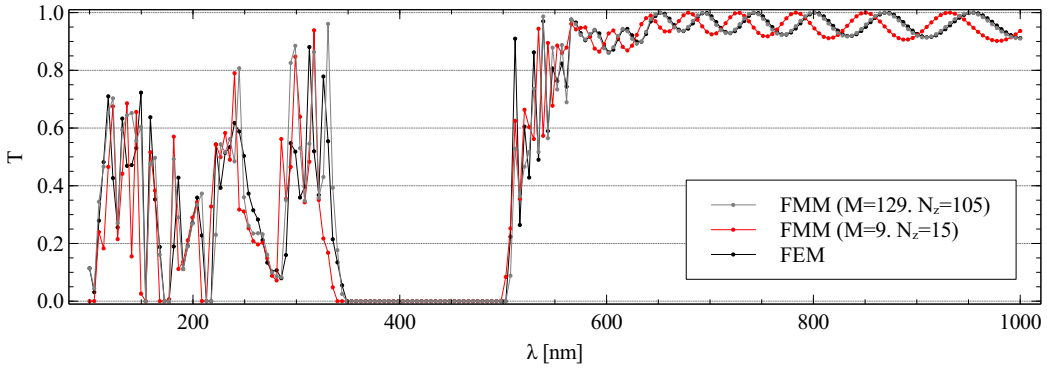
(a) FEM grid (b) Staircase approximation

Figure IV.13: *Photonic Crystal slab in 2D adapted from PhC waveguide of [4]. FEM discretization of the PhC consisting of background material (green) with $\varepsilon = 12$ and vacuum rods (grey) is displayed in (a). In contrast to all other structures in this study, illumination is from the left. Three unit cells are plotted above one another. Staircase approximation with only $N_z = 5$ unique layers (b) leads to 107 layers in total (cf. main text).*

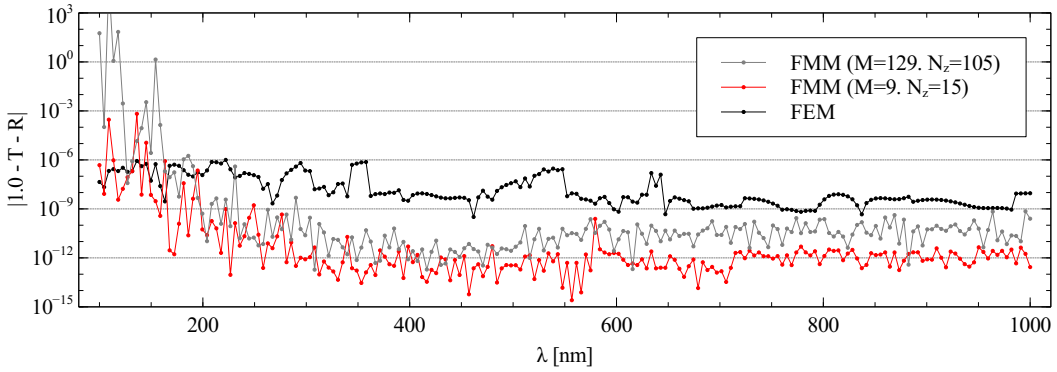
Since we do not intend to study supercells for aperiodic structures here, we use a simple slab with no missing rows. Additionally, we make use of the fact that several layers are repeated from left to right [Fig. IV.13(b)]. In order to reduce numerical effort, the layer-eigenproblem of FMM is only solved for nonequal layers. This number of unique layers is denoted by N_z. For example, for $N_z = 5$ unique layers the PhC slab consists of 107 layers in total. For all of these layers, the scattering matrix algorithm has to be performed but not finding the solution of the eigenvalue problem. The latter only has to be solved N_z times. This is a standard option of S^4 [50].

Figure IV.14(a) shows a bandgap in the transmission spectrum of the PhC slab from approximately 350 nm to 500 nm. FEM results are converged to more than six digits on average of all illumination wavelengths λ, where illumination is perpendicular to the left boundary of the CoDo. Already for a very small number of Fourier harmonics $M = 9$ and $N_z = 15$ unique layers, FMM results show the general trend of the transmission spectrum. However, a significant shift of the cut-off frequency to approximately 340 nm is obtained. Additionally, the Fabry-Pérot-like spectrum for long wavelengths $\lambda > 650$ nm is blue-shifted as well. Here, the shift increases to more than 20 nm.

The spectral shift of the transmission computed with the FMM is partly due to a small number of Fourier harmonics M, but also caused by an inaccurate staircase approximation, i.e. the number of layers. So FMM is limited by two independent numerical parameters when using staircasing: the size of the basis set M and the geometric approximation via the number of layers (cf. N_z). To obtain a similar transmission spectrum as FEM results, we increase both parameters to $M = 129$ and $N_z = 105$, i.e. 1873 layers. Then the transmission spectra of the FEM and FMM are comparable.



(a) Transmittance spectrum



(b) Energy conservation

Figure IV.14: *Transmittance spectrum (a) of 2D PhC slab. Staircasing introduces a spectral (blue-) shift to the transmission spectrum for a low number of layers (red line). Here, $N_z = 15$ unique layers, i.e. 285 layers in total, are computed with the FMM. Using an option of S^4, only the eigenvalue problems of the unique layers are solved, while for all layers the scattering matrix algorithm is performed. The shift can be reduced by both using more Fourier harmonics M and layers (grey line). However, the interplay of both numerical parameters is unpredictable and simply increasing one does not yield more accurate results in general. FEM simulations (black line) show a bandgap from approximately 350 nm to 500 nm. Analysing energy conservation (b) leads to the conclusion that FEM simulations fulfil energy conservation over the whole spectral range. FMM leads to faulty reflection spectra for short wavelengths $\lambda < 150$ nm, yet very good agreement for greater λ.*

On the other hand, the reflection spectrum of FMM simulations is corrupted: When analysing energy conservation [Fig. IV.14(b)], violation of energy conservation is obtained for small wavelengths $\lambda < 150$ nm. This contradicts the prediction of Moharam that FMM satisfies energy conservation in all circumstances [56]. For the long wavelength range, however, this is particularly true: energy conservation is fulfilled to an accuracy of less than 10^{-9} for FMM simulations. FEM results show accuracy of less than 10^{-6} uniformly for the illumination spectrum of $100 \leq \lambda \leq 1000$ nm.

Nevertheless, the accurate results of energy conservation of the FMM disguise erroneous results of the complex Fourier coefficients. Again, we analyse the error of the absolute value of the FT ($\Delta_A^{(f)}$) and the phase included error ($\Delta_F^{(f)}$) for the forward propagating modes in Figure IV.15. Errors for short wavelengths are significantly higher than in the transmission spectrum since faulty high diffraction orders contribute to the average error (see Sec. IV.2.2). For long wavelengths, for which only one diffraction order exists, we see similar results to the previous discussions. The absolute value is well approximated (errors of less than 10^{-2}), but the correct phase is only computed to an accuracy of approximately 10%. Furthermore, the spectral shift yields oscillatoric pattern in the $\Delta_A^{(f)}$ error.

This phase error could not be scaled down by using more Fourier harmonics M. Rather, the interplay of M and the number of layers needs to be analysed carefully when aiming for very accurate results using the FMM. Once again, we conclude that the FMM is suitable for energy observations such as band diagrams of PhCs, but has to be handled carefully when resolving the total far-field pattern, where accurate complex Fourier coefficients are needed.

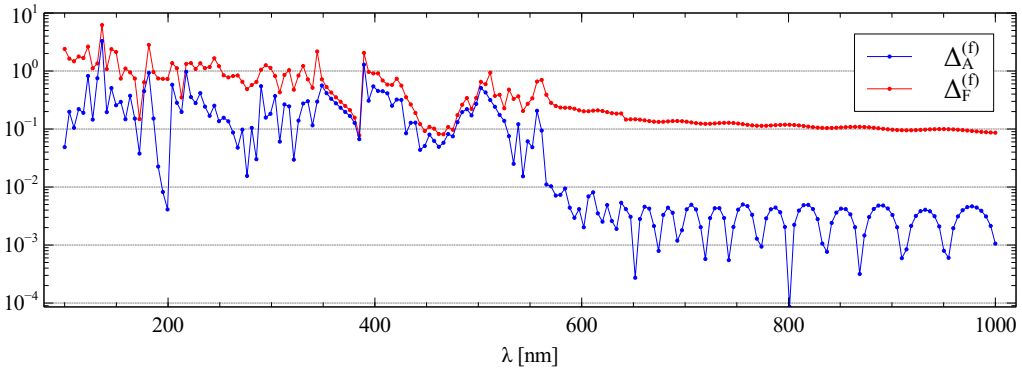


Figure IV.15: *Errors of FMM simulation ($M = 129$, $N_z = 105$) for the forward propagating Fourier coefficients. The good agreement of the transmission spectra of FEM and FMM (Fig. IV.14) is confirmed by small errors of the absolute values ($\Delta_A^{(f)}$, blue line). However, the phase correlations of the Fourier coefficients is not the same as the converged FEM results even for long wavelengths λ ($\Delta_F^{(f)}$, red line). This is due to both, limited accuracy of the basis because of the chosen number of Fourer harmonics M and the limiting approximation of the geometry because of staircasing. The latter errors accumulate for these high number of 1873 layers (cf. discussion of Figure IV.8).*

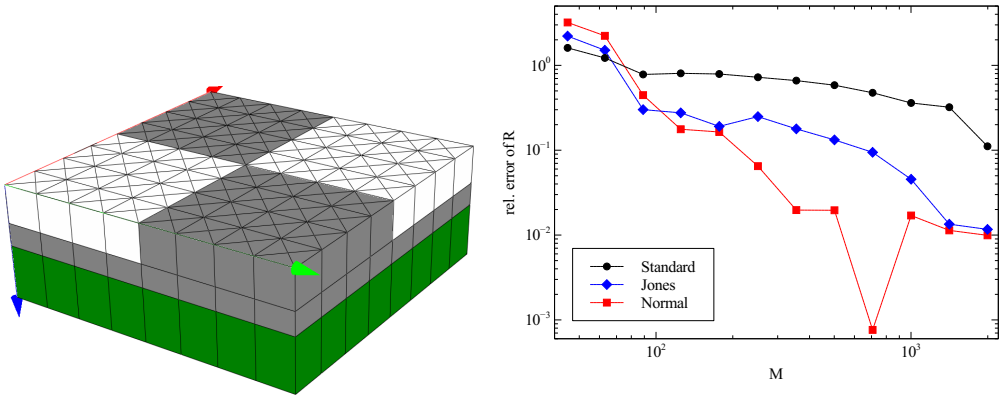
IV.4 3D Simulations

Due to the progress in both numerics and computing power, modelling of 3D problems has become much more important in the field of nano-optical scattering problems. However, rigorous simulations of Maxwell's equations on a nanometre scale require either long computing time (e.g. time-domain methods like FDTD) or high memory consumption (e.g. frequency-domain methods such as FMM and FEM). That is why convergence behaviour of the different methods used for nano-optical simulations is of great interest in order to be able to classify numerical methods with respect to their requirement of resources. In the following, we are more interested in general convergence characteristics of the FMM to analyse the applicability of this method, which was initially formulated for 2D grating problems, to 3D problems.

We start by analysing the extension of simple lamellar gratings to so-called checkerboard gratings with absorbing material in Section IV.4.1. Additionally, we compare the different formulations of the FMM (see Sec. III.1.4) for a quadratic pin hole in an absorbing layer. These kinds of structures are important for lithography processes in the semiconductor industry. Finally, we give a qualitative comparison of FEM and FMM results for the simplified band diagram of a dielectric three-dimensional PhC slab.

IV.4.1 Checkerboard Grating

Simple lamellar gratings are periodic in only one (the x-) direction. A simple version of twofold periodic gratings is a checkerboard grating made of quadratic boxes [Fig. IV.16(a)]. We use a grating with pitch $p_x = p_y = 400$ nm and side length of the boxes $w_x = w_y = 200$ nm. These boxes are made out of a material with refractive index $n_2 = 4.294 + 0.044165i$ (dark grey domain) and are 50 nm in height. An additional slab, which is 30 nm high, is placed below the grating and has refractive index n_2 as well. The structure is placed on a substrate with refractive index $n_3 = 1.45$ (green domain). A plane wave with wavelength $\lambda_0 = 500$ nm and incident spherical angles $\theta = 30°$ and $\phi = 10°$ illuminates the grating from vacuum ($n_1 = 1.0$, light grey domain). We compare both s-polarization (TE) and p-polarization (TM).



(a) Geometric setup

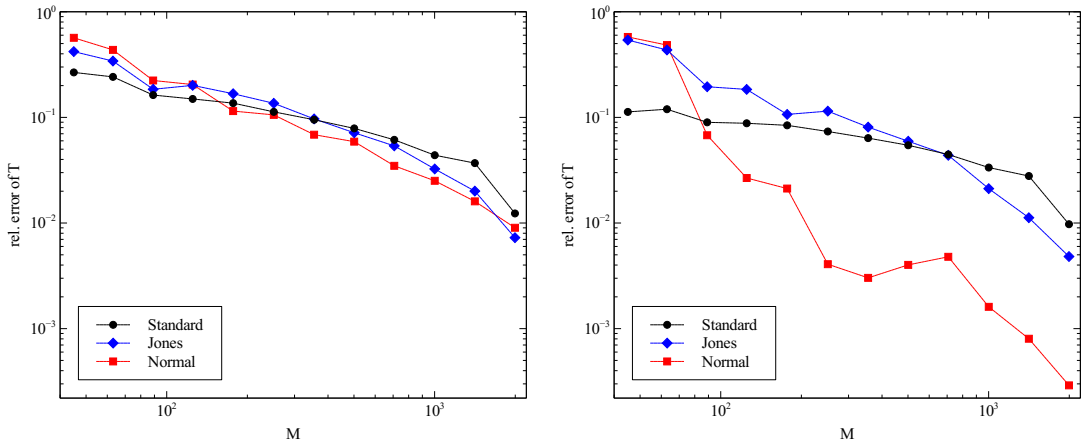
(b) Convergence of $R(s)$

Figure IV.16: *Checkerboard grating with quadratic footprint (a). The boxes and a slab below are made out of absorbing material (dark grey domain). The grating is placed on a dielectric substrate (green domain) and illuminated from above (vacuum, light grey domain) in conical mount (cf. to main text for refractive indices, wavelength and spherical angles). Convergence of the different formulations of the FMM is analysed with respect to each best result for $M = 2819$ Fourier harmonics (b). Here, the figure of merit is the reflectance R for s-polarized illumination. Standard FMM (black circles) shows slow convergence due to inappropriate FFR. Decomposition of the electric field into normal and tangential components yields faster convergence for the normal vector method (red squares) compared to the Jones vector field basis (blue diamonds).*

First, we investigate convergence for s-polarization of the different formulations of the FMM (see Sec. III.1.4) with respect to their best results ($M = 2819$), respectively [Fig. IV.16(b)]. Here, we analyse the relative error of the reflectance R, i.e. the total reflected energy flux in the z-direction (see Sec. II.2.3). Due to symmetry reasons only the 0-th diffraction order contributes to R. The vanishing higher diffraction orders are well approximated by the FMM. For this grating, we confirm general findings of the convergence of the FMM [49]: standard FMM, i.e. not applying proper FFR in 2D layers, yields slow convergence. Using a Jones vector field to decompose normal and tangential components of the electric field results in much faster convergence down to a relative error of 10^{-2}. When obtaining the vector field with the so-called normal vector method, the convergence rate is much higher. However, it becomes non-monotonous (cf. trough at $M \approx 700$).

Although the Jones and the normal vector method show comparable relative errors, the values of reflectance for s-polarization [$R(s)$] for the best results differ in more than 10% (Tab. IV.1). Reflectance obtained with standard FMM is several orders different from advanced FMM and FEM simulations. Due to the missing convergence theory of the FMM it is *a priori* not clear which result is the best approximation. In particular, this is of great interest since usually only hundreds of harmonics are used for FMM computations [49] and it cannot generally be confirmed whether all variants converge to the same result.

Furthermore, error bounds and general convergence behaviour depend strongly on the specific problem: in Figure IV.17 we compare convergence of the FMM for p-polarized [Fig. IV.17(a)] and s-polarized [Fig. IV.17(b)] illumination, respectively. The relative error of the transmittance T with respect to the best solution ($M = 2819$) is again plotted for each FMM formulation separately. Illumination in p-polarization yields in general more inaccurate results [note equal y-axis scaling of (a) and (b)]. Convergence characteristics of the standard FMM are comparable as expected, since here suitable and unsuitable Fourier Factorization Rules are applied for both s- and p-polarization.



(a) p-polarization (b) s-polarization

Figure IV.17: *Convergence of the transmittance T for p-polarized (a) and s-polarized (b) illumination with respect to each best result ($M = 2819$) for three different formulations of the FMM (see Sec. III.1.4). Convergence characteristic for p-polarization is comparable for all variants. However, values of T differ by more than 10% (Tab. IV.1) and from FMM convergence theory it is not clear which one is correct. Results for s-polarization confirm findings of [49] that the normal vector method (red squares) converges faster compared to the Jones vector field basis (blue diamonds). In comparison to FEM results, standard FMM (black circles) yields inaccurate results.*

On the other hand, decomposition into normal and tangential components with the help of the Jones and the normal vector method leads to much better convergence for s-polarized illumination [Fig. IV.17(b)], but not for p-polarization. For the latter convergence patterns of all three investigated formulations are comparable. Nevertheless, the exact values for transmittance $T(p)$ differ by more than 10% again (Tab. IV.1). Assuming that FEM results are the best approximation (see Sec. III.2.3), the normal vector method yields the best results for R, T and the 0-th order reflection and transmission coefficients, respectively.

Interestingly, the convergence of the sum of the reflectance and transmittance for s-polarization, i.e. the absorption $A(s)$, converges quite differently compared to the FEM results (Fig. IV.18): in general, errors are higher than comparing errors to the best FMM results [note different y-axis scaling in Fig. IV.18 compared to Fig. IV.17]. Additionally, in contrast to FMM-compared convergence, the Jones vector method yields well approximated absorption coefficients for some numbers of Fourier harmonics. For $M \approx 190$ the relative error with respect to FEM results decreases down to 2×10^{-3}. However, increasing M yields worse results. This potentially originates from different absorption coefficients obtained with FMM and FEM even for $M \to \infty$.

Again, since FEM solutions are only bounded by the regularity of the analytic solution, we expect them to approximate the values correctly. The Fourier approximation of the FMM does not seem to lead to results which are equal to FEM simulations to more than one digit for this checkerboard grating. However, if only results which are exact up to 10% are needed, the FMM gives insight into the physics of the device in a small amount of computing time. Note that the recommended option for S^4 is the normal vector method [50]. In spite of this practical experience, the Jones vector method yields (in certain circumstances) results for A which are closer to those of FEM simulations. Additionally, the equivalence of normal and Jones vector method results for the absorption originates from lower and higher R and T values, respectively. It should be further analysed if both methods fulfil an energy conservation constraint which leads to this overlap for A.

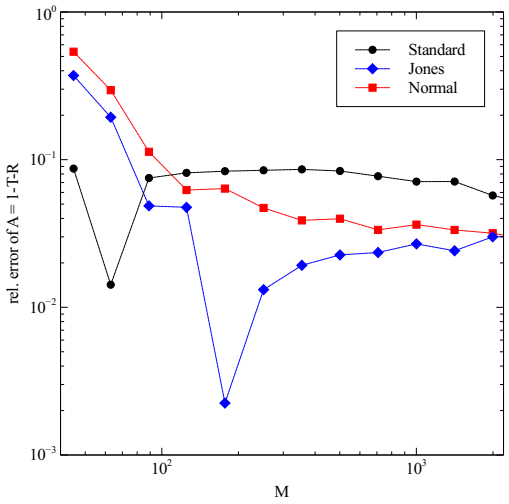


Figure IV.18: *Convergence with respect to FEM results of the absorption A for s-polarized illumination obtained with three different FMM formulations. In contrast to Figure IV.17, the Jones vector field method (blue diamonds) shows smallest relative errors, especially for $M \approx 190$. Standard FMM (black circles) yields inaccurate results and convergence of the normal vector method (red squares) is monotonous but does not agree to more than 3×10^{-2} with the FEM value.*

Table IV.1: *Best results for FEM (more than 10^6 unknowns) and FMM ($M = 2819$) simulations. Agreement between FEM and the normal vector method (Normal) and the Jones vector field basis (Jones) are obtained for absorption with p-polarized illumination $A(p)$. However, values for reflectance R, transmittance T and their 0-th order values R_0, T_0 differ. In particular, standard FMM (Std.) shows high discrepancy for $R(s)$. Due to symmetry reasons, higher order reflection coefficients vanish, i.e. $R \approx R_0$, which is well approximated by the FMM.*

	FEM	Std.	Normal	Jones
$T(s)$	0.458	0.493	0.449	0.428
$T(p)$	0.396	0.429	0.378	0.366
$R(s)$	0.126	0.070	0.147	0.169
$R(p)$	0.155	0.102	0.175	0.189
$T_0(s)$	0.431	0.467	0.421	0.403
$T_0(p)$	0.271	0.309	0.255	0.245
$R_0(s)$	0.126	0.070	0.147	0.169
$R_0(p)$	0.155	0.102	0.175	0.189
$A(s)$	0.416	0.437	0.404	0.403
$A(p)$	0.449	0.468	0.447	0.445

IV.4.2 Pin Hole

In this section we investigate again a twofold periodic quadratic structure: an absorbing layer with refractive index $n_2 = 2.343 + 0.586i$ is placed on top of a substrate with $n_3 = 1.563$. A so-called quadratic pin hole with side length 300 nm and refractive index $n_1 = 1.0$ is left in the absorbing layer. The pitch in both the x- and the y-direction is 800 nm. The structure is illuminated from below with a perpendicularly propagating plane wave with $\lambda_0 = 193$ nm. The incident wave is polarized in the y-direction and the superstrate is vacuum. The corresponding field distribution shows singularities at the edges of the hole and in particular at the upper boundary of the hole (Fig. IV.19). That is why in the FEM grid, extra edges in all $\pm y$-, $\pm x$- and $\pm z$-directions are added at a distance of 15nm from the edges of the hole. The initial FEM mesh is obtained with a side length constraint of 0.7λ in the substrate and the hole region with the material dependent wavelength $\lambda = \lambda_0/n_i$, respectively.

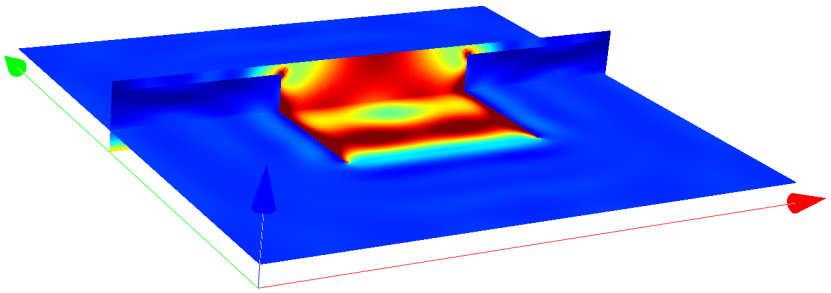


Figure IV.19: *Intensity distribution of quadratic pin hole in absorbing layer on dielectric substrate. Side length of the hole is 300 nm and the quadratic unit cell is 800 nm wide. Incident plane wave with wavelength $\lambda_0 = 139$ nm propagates upwards. The field distribution of FEM simulation ($p = 4$) reveals singularities at the edges and especially in the upper part of the hole. These are resolved with a manually locally refined grid.*

The two-dimensional triangular mesh is extruded in the z-direction with the meshing tool of *JCMsuite*. In doing so, a grid composed of prisms is obtained. Due to faster convergence rate for the polynomial degree (see Sec. III.2.3), convergence of FEM is analysed with this grid which is optimized for singularities. The relative maximal Fourier error ($\Delta^{(i)}_{F,\infty}$, Def. II.4.8) shows accurate p-convergence with respect to the best numerical result for $p = 7$ (Fig. IV.20). However, it should be noted that higher diffraction orders [e.g. $(N_1, N_2) = (\pm 1, \pm 4)$] are limited to $\Delta^{(N_1,N_2)}_{F,\infty} \approx 10^{-2}$. These slowly converging Fourier coefficients originate from the field singularities and illustrate the irregularity of the electromagnetic fields for this structure.

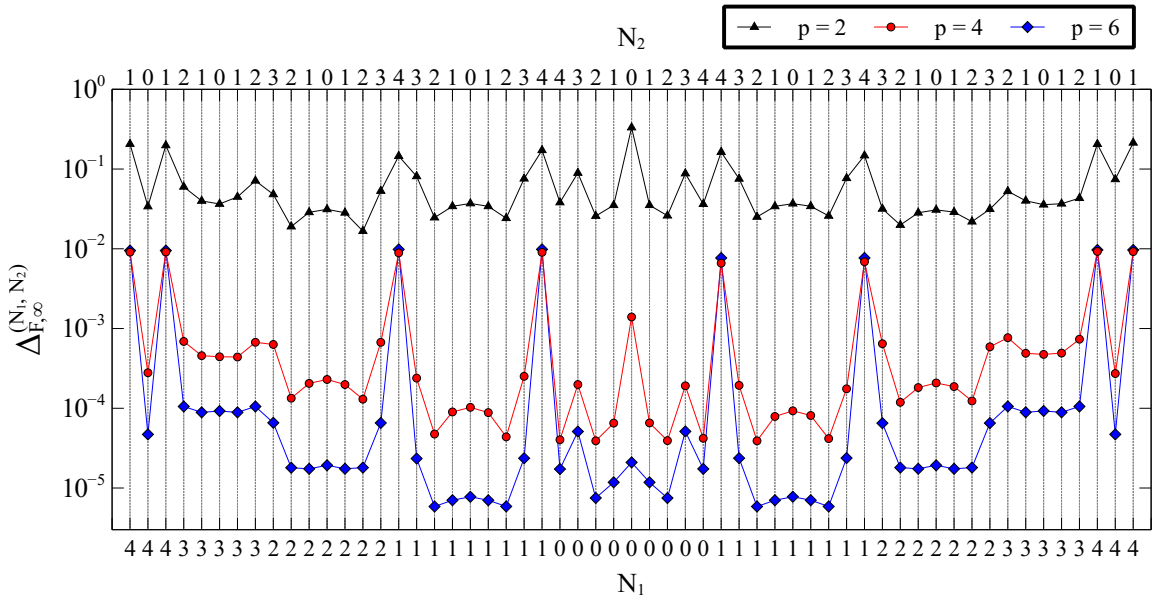
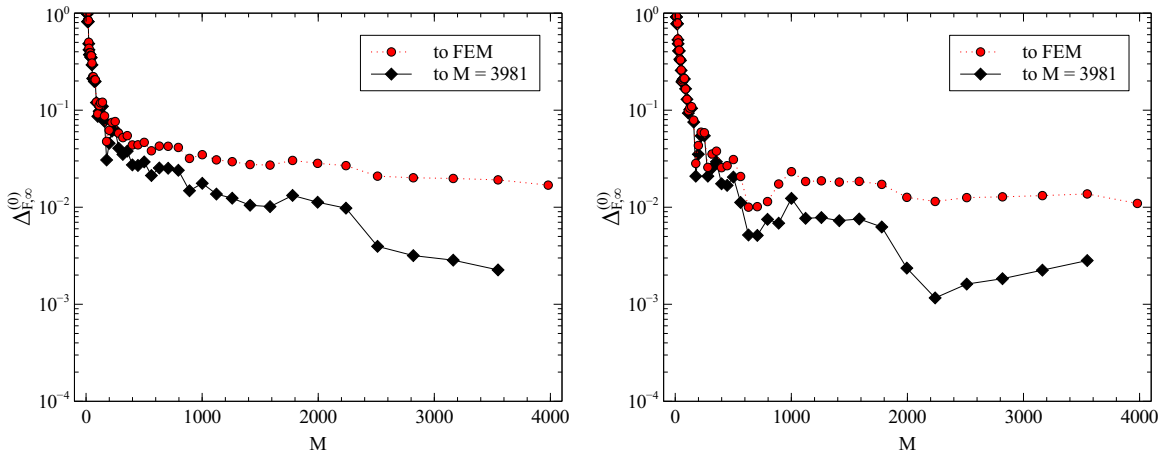


Figure IV.20: *Convergence of diffraction orders of FEM results. Diffraction order in the x-direction (N_1) is displayed on the lower x-axis and diffraction order in the y-direction (N_2) on the x-axis at the top. Both are sorted in ascending order, i.e. numbers left of 0 are negative diffraction orders and numbers right of 0 are positive diffraction orders [e.g. the outermost left order is $(N_1, N_2) = -4, -1$]. Convergence of the relative maximal Fourier error ($\Delta^{(i)}_{F,\infty}$, Def. II.4.8) is plotted for different polynomial degrees p with respect to the results for $p = 7$. The 0-th order Fourier coefficient converges well, while errors of higher diffraction orders [e.g. $(N_1, N_2) = (\pm 1, \pm 4)$] saturate.*

In order not to analyse these limiting high diffraction orders, we compare the relative errors of the FMM for the 0-th order Fourier coefficient (Fig. IV.21). First, we investigate the standard FMM and the normal vector method with respect to their respective best FMM results ($M = 3981$) and with respect to the FEM value for $p = 7$. Again, the standard formulation of the FMM, which dismisses the appropriate application of the Inverse Rule, converges slowly and only to 2×10^{-2} relative to the 0-th order Fourier coefficient obtained with the FEM [Fig. IV.21(a)]. Furthermore, usual oscillatoric convergence characteristics of small numbers of Fourier harmonics ($M \lesssim 600$) can be seen. The error compared to the best FMM result overestimates the error compared to the FEM result by roughly one order of magnitude.

On the other hand, the normal vector method [Fig. IV.21(b)] yields accurate results much faster: Already for $M \approx 600$, the value of the 0-th order Fourier coefficient differs only 1% from the FEM result for the relative maximal Fourier error. This confirms the widespread experience to use a number of harmonics in the mid-hundreds [49]. However afterwards, this specific error increases for $M \to 1000$ in both FMM and FEM error measure. Additionally, the error decreases significantly for the FMM comparison for $M \approx 2200$ and increases again. This unpredictable convergence behaviour of the FMM causes problems in the judgement whether a numerical result of this method is converged or not.

In contrast to the checkerboard grating analysed in the previous section, the FMM extended with a Jones vector field yields better convergence than the normal vector method [Fig. IV.22(a)]: for $M = 4000$, the 0-th order Fourier coefficient equals approximately 7×10^{-3} to the FEM simulation. The self-consistent FMM convergence overestimates the relative error again by roughly one order of magnitude. Yet convergence behaviour of the Jones basis FMM is much smoother compared to the normal vector method. Excluding peaks at $M \approx 2000$, convergence is monotonous and inherent oscillatoric effects of the FMM are flattened.



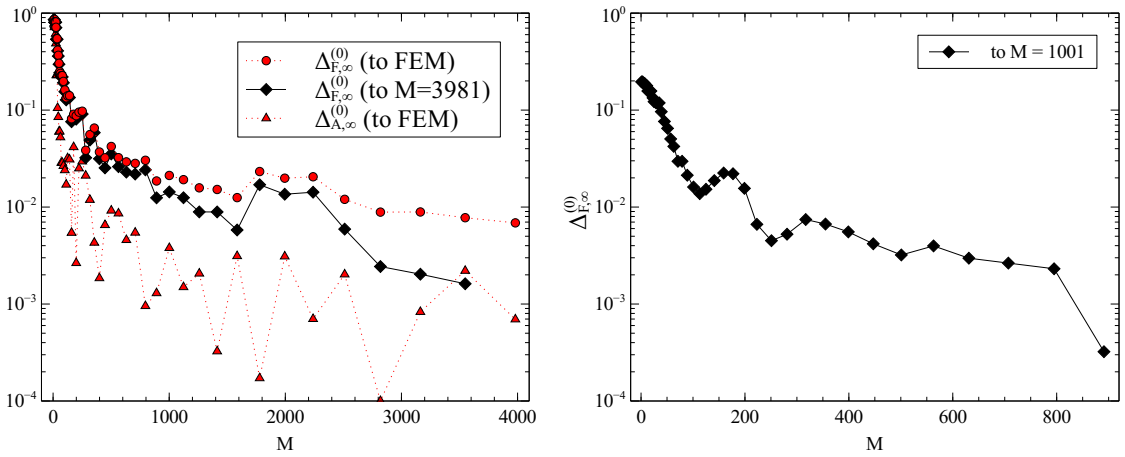
(a) FMM Standard

(b) FMM Normal

Figure IV.21: *Convergence of the 0-th order Fourier coefficient for standard FMM (a) and FMM with the normal vector method (b) is compared to the best FEM result ($p = 7$, red circles) and each best FMM result ($M = 3981$, black diamonds). Standard FMM shows slow convergence and self-consistent FMM computation of the relative error overestimates accuracy in roughly one order of magnitude. Agreement up to 2% is obtained while normal vector FMM agrees approximately 1% with the FEM result. Convergence rate of the later is much higher, particularly for $M \lesssim 600$. However, high fluctuations in the error limits the ability to judge whether FMM results are already converged.*

In order to verify findings of well approximated propagating energy in 2D problems (cf. previous sections), we compare the error of the complex Fourier coefficient to the one of the absolute value $\Delta^{(0)}_{A,\infty}$ [Fig. IV.22(a)]. We see that already for $M \approx 500$ the error of the energy propagated by the 0-th diffraction order stabilizes below an upper bound of 1%. The sinusoidal convergence of the FMM, which is reported throughout literature, is clearly visible and leads to oscillations of more than one order of magnitude. Although relative errors of less than 10^{-4} are obtained with the Jones vector field formulation of the FMM, we are once again facing the problem that the peaks in the convergence inhibit the possibility to estimate how accurately a result can be computed when only using the FMM.

We also analyse the performance of the FMM including subpixel smoothing [Fig. IV.22(b)]. Convergence with respect to the best ($M = 1001$) result obtained using this method shows a fast convergence rate and errors of less than 10^{-3}. However, the value of the 0-th Fourier coefficient for the best subpixel averaging result differs significantly from those of the standard FMM, the normal vector FMM, the Jones basis FMM and FEM simulations. For instance, FEM simulations yield $f^{(0)}_{\mathrm{FEM}} = 0.1075 - 0.0357i$ at the upper boundary of the chosen computational domain. The result of FMM including subpixel averaging is $f^{(0)}_{\mathrm{Subpixel}} = 0.2500 - 0.0436i$ for $M = 1001$ which is obviously different to the other results. This seems to be introduced by a limiting resolution of the FFT used to obtain the subpixel averaged anisotropic permittivity tensor. The oversampling factor was set to eight, which should generally be sufficient. Nevertheless, convergence behaviour of subpixel smoothing might be improved by using higher oversampling factors which on the other hand lead to higher numerical effort.



(a) FMM Jones

(b) FMM Subpixel FFT

Figure IV.22: *Convergence of the 0-th order Fourier coefficient for FMM with Jones vector field basis (a) and FMM with subpixel averaging (b) is compared to the best FEM result ($p = 7$, red circles) and each best FMM result (black diamonds). The Jones vector field method does not yield faster convergence compared to the normal vector method for small numbers of Fourier harmonics M [Fig. IV.21(b)]. However, convergence characteristic is smoother and the best result shows a relative maximal Fourier error of less than 1%. Again, the energy propagating in the 0-th diffraction order ($\Delta^{(0)}_{A,\infty}$) shows much smaller errors down to 10^{-4}. Nevertheless, oscillatoric convergence typical for the FMM is observed. Although subpixel averaging shows small errors with respect to its best FMM result ($M = 1001$), the value of the 0-th Fourier coefficient differs in roughly a factor of two from FEM and other FMM results (cf. main text for details).*

IV.4.3 Photonic Crystal

In this section we analyse a dielectric Photonic Crystal slab in 3D. First of all, it should be noted that for obtaining band diagrams from solving Maxwell's equations, resonance problems are in general more suitable than the formulation as a scattering problem. However, nano-optical scattering problems are often solved to compute reflectance and transmittance spectra of PhC structures [4, 50]. Therefore, we compare FMM to FEM results for a simplification of [8]: the device depicted in Figure IV.23 is illuminated from vacuum ($n_1 = 1.0$, grey domain). It consists of hexagonally arranged vaccum holes with radius $r = 210$ nm in silicon (n_2, red domain). The slab is placed on top of a glass substrate ($n_3 = 1.53$, green domain). The hexagonal pitch is 300 nm and the slab is 195 nm high. First of all, compared to [8], we neglect the side wall angle (17°) of the holes which is determined by the crystal structure of silicon. Furthermore, we use the dispersion relation of silicon [60] yet we only use $\mathrm{Re}[n_{\mathrm{Si}}(\omega)]$ for n_2. These two simplifications lead to a more suitable problem for the FMM, since we do not need any staircasing in the z-direction for the side wall angle and the questionable completeness of the plane wave basis for complex permittivity profiles does not need to be taken into account.

We perform a wavelength scan for $562.5 \leq \lambda \leq 1012.5$ nm with incident angles $10° \leq \theta \leq 80°$ (rotation about the y-axis). The illuminating plane wave is polarized in the x-direction (red axis in Fig. IV.23). For FEM simulations we use an extruded geometry which consists of prisms oriented in the z-direction. Slicing of these prisms in the z-direction as well as PML discretization is automatically adjusted for each λ and θ by *JCMsuite*. For these purposes the numerical parameter *PrecisionFieldEnergy* is set to 10^{-3}. The finite element polynomial degree p is set to 2 which yields relative errors of reflectance of less than 10^{-3} with respect to $p = 3$. This comparison is performed on a coarse equidistant ($\lambda \times \theta = 9 \times 9$)-grid of the wavelength and angle intervals mentioned above. Reflectance R as the figure of merit is plotted on a 2D colour plot in Figure IV.24(a). A complex structure of bands where $R \approx 1$ is obtained. However, it should be noted that the equidistant spacing of 129 wavelengths and 65 angles might be too large for analysing the fundamental physics of the slab (e.g. a band starting at $\theta = 10°, \lambda \approx 890$ nm is not clearly visible). Nevertheless, this setup yields diagrams which can be compared for the FMM and the FEM.

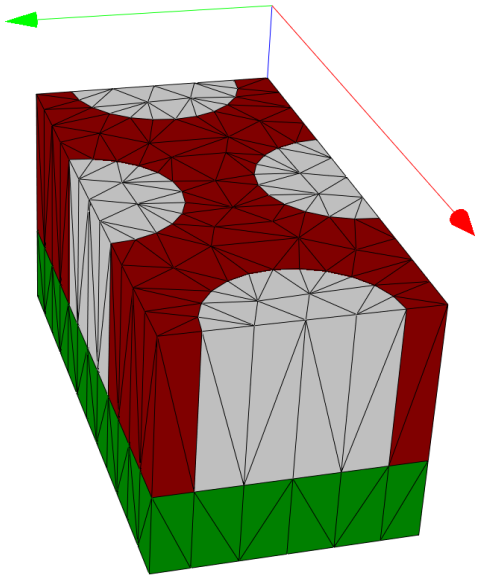


Figure IV.23: *PhC slab consisting of vacuum holes (grey domain) in silicon (red domain) placed on top of a glass substrate (green domain). The device is illuminated from above with a plane wave of wavelength λ and incident angle θ which represents a rotation about the y-axis (green axis). The wave is polarized in the x-direction (red axis). For FEM simulations also a hexagonal unit cell would be possible [8].*

Table IV.2: *Simulation times and energy conservation errors of FEM, FMM with the normal vector method (Normal) and FMM with a Jones polarization basis (Jones), respectively. The energy conservation error* $\Delta := \log_{10}(|1.0-T-R|)$ *is analysed with respect to its mean and maximal value, its standard deviation* σ *and the amount of data points which violate energy conservation by more than one per thousand* ($\Delta \geq -3$) *for two numbers of Fourier harmonics* $M = 99, 499$. *FEM simulations are parallelized on two CPUs, whereas FMM is performed only on one CPU (here, total and CPU time equal one another). Using a Jones vector polarization basis requires more computation time, yet yields more accurate results compared to the normal vector method. Additionally, errors are more concentrated (cf.* $\sigma < 0.5$*). Shifts of the resonance frequencies (Fig. IV.24) are not controlled by this error analysis and are challenging for both methods because of small widths of these peaks.*

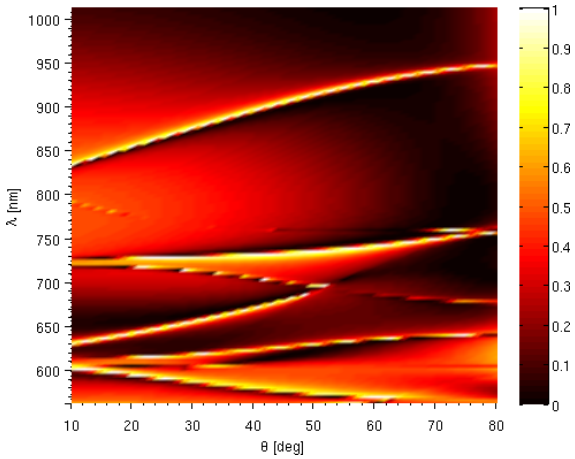
	FEM	Normal		Jones	
M		99	499	99	499
mean(Δ)	-3.994	-3.588	-4.479	-3.481	-3.957
max(Δ)	-1.223	-0.685	-1.386	-1.262	-1.724
$\sigma(\Delta)$	0.601	0.813	0.851	0.496	0.403
$\Delta \geq -3$ [%]	3.86	26.98	4.22	15.41	3.45
total [10^4s]	134.5	1.5	216.1	1.7	301.9
CPU [10^4s]	235.1	1.5	216.1	1.7	301.9

As previous results (see Sec. IV.3.2) suggest, we expect a frequency shift for increasing numerical effort for both the FEM (increasing polynomial degree p) and the FMM (increasing number of Fourier harmonics M). That is why in Table IV.2 we analyse the energy conservation error $\Delta := \log_{10}(|1.0 - T - R|)$ of the reflectance R and the transmittance T, rather than comparing each data point of FMM simulations with those of FEM results. The mean energy error of the FEM results is approximately 10^{-4} and is more or less concentrated with a standard deviation $\sigma = 0.601$. Furthermore, we checked how many data points fulfil energy conservation by less than one per thousand ($\Delta \geq -3$). For FEM results, these are only 4%. We used parallelization with two CPUs yielding a total simulation time of 374 hours which is distributed on a cluster of 192 available CPUs [36]. For the FMM, no parallelization is enabled and, accordingly, total and CPU time in Table IV.2 are equal.

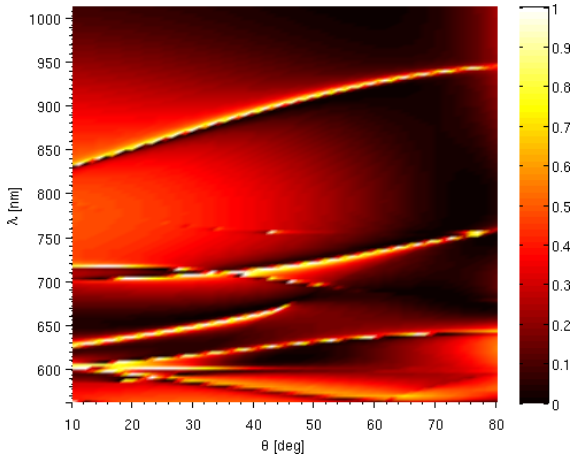
Due to the findings of the previous sections, we analyse the FMM extended by the normal vector method [Fig. IV.24(b) and IV.24(c)] as well as the Jones polarization basis [Fig. IV.24(d) and IV.24(e)] for $M = 99$ and $M = 499$ Fourier harmonics, respectively. The normal vector method yields fast results, yet the energy conservation error is largely distributed ($\sigma > 0.8$, Tab. IV.2) and energy conservation is violated maximal by more than 10% for $M = 99$. This originates from values $R+T > 1$ for certain data points which again contradicts Moharam's prediction that the FMM fulfils invariably the conservation law of energy. However, the pattern of the band diagram obtained with $M = 99$ Fourier harmonics gives a first impression of the band structure of the slab. In particular, the separated bands starting at $\theta = 10°$ and $\lambda \approx 720$ nm and $\lambda \approx 730$ nm respectively [Fig. IV.24(a)], are not well resolved and an additional crossing at $\theta \approx 35°, \lambda \approx 710$ nm is introduced by numerical inaccuracies [Fig. IV.24(b)]. The pattern for $M = 499$ is comparable to FEM results [Fig. IV.24(c)], however, frequency shifts lead to slightly different behaviour particularly for large incident angles.

The FMM with a Jones polarization basis requires more computation time (10% for $M = 99$ and 40% for $M = 499$) compared to the normal vector method. Yet, results fulfil energy conservation much better (maximal errors are smaller than $10^{-1.2}$) and are much more concentrated ($\sigma = 0.403$ for $M = 499$). Although CPU time is higher than the one for FEM, energy investigations suggest that results are more accurate since only 3.5% of the data points violate energy conservation by more than one per thousand. Furthermore, plots of the reflectance are comparable to FEM simulations for both analysed numbers of Fourier harmonics [Fig. IV.24(d) and IV.24(e)].

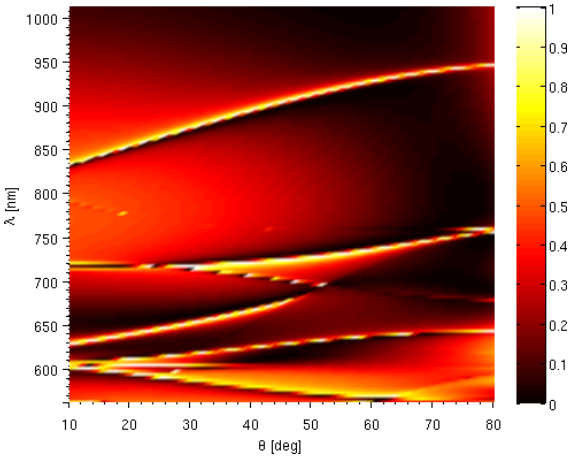
In conclusion, we found that results of the FMM and the FEM are generally comparable for this dielectric PhC slab. We confirm findings of Antos [2] that the FMM with a Jones polarization basis is well suited for circular (and probably also elliptic) geometrical features, whereas the normal vector method yields less accurate results for the same number of Fourier harmonics. Furthermore, shifts of resonance frequencies are challenging for both the FMM and the FEM. However, FEM offers generally more numerical parameters (e.g. polynomial degree p, grid discretization, PML discretization) to optimize convergence of these resonance peaks in the reflectance spectrum. Again, we found that all variants of the FMM introduce erroneous phase shifts in the Fourier coefficients, but lead to accurate results for the analysis of energy.



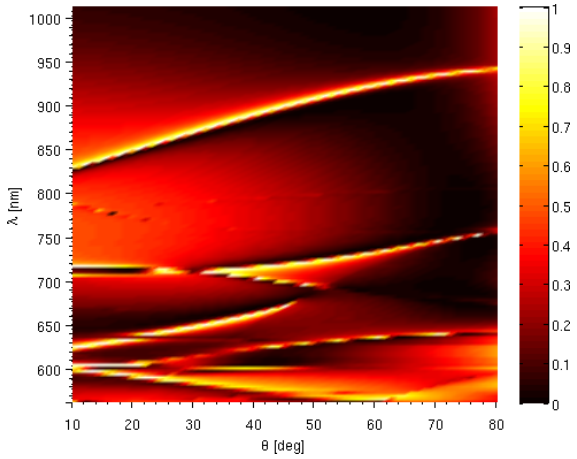
(a) FEM



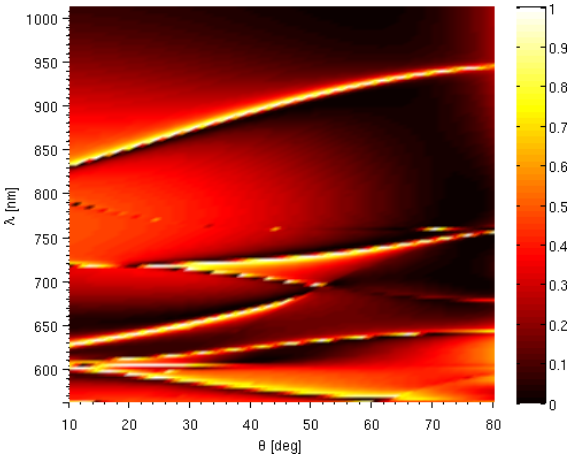
(b) FMM Normal ($M = 99$)



(c) FMM Normal ($M = 499$)



(d) FMM Jones ($M = 99$)



(e) FMM Jones ($M = 499$)

Figure IV.24: *Band diagrams of PhC slab obtained with the FEM, the normal vector method FMM (Normal) and the FMM with a Jones vector field (Jones). FEM results are computed using polynomial degree $p = 2$ and discretization in the z-direction as well as the exterior PML discretization is controlled with the numerical parameter PrecisionFieldEnergy of JCMsuite with a value of 10^{-3} (cf. main text for details). We use an equidistant parameter grid with 129 wavelengths λ and 65 incidence angles θ. The reflectance R is plotted on a colour scaling. The normal vector method with $M = 99$ introduces unexpected crossings and unseparated bands, yet it yields comparable results for $M = 499$. Jones polarization method's results already match for $M = 99$. For these settings FMM and FEM yield similar band structures which is expected from findings of previous sections that erroneous phase shifts introduced by the FMM do not limit its applicability to the analysis of energy scattering.*

Chapter V

Summary

The aim of this convergence study was to investigate general convergence characteristics of the Fourier Modal Method for nano-optical scattering problems. Due to the non-existence of a rigorous convergence theory for this method, we used FEM simulations as reference solutions. To compute results of the FMM we used a modified version of the open-source solver S^4 [70] and FEM simulations were performed with the software package *JCMsuite* [32]. For systematic comparison we unified the interface of both solvers.

We started by stating Maxwell's equations as a common model for nano-optical scattering problems. In particular, we noted electromagnetic field properties of periodic structures, since the FMM is inherently periodic in the x-direction for 2D problems and twofold periodic in 3D. In the 1980s, FMM originated from grating theory. That is why we briefly described its ideas and showed that in this field the diffraction efficiencies, i.e. the absolute values of Fourier coefficients, are of interest. Accordingly, we defined different investigated errors and remarked that the error of absolute values for both the Fourier coefficients and the near-field generally underestimates the errors including phase shifts. The latter are of great interest in modern physics, e.g. metrology and plasmonics.

First known as RCWA [54], the term Fourier Modal Method was coined by Li in 1996. He mathematically justified the major breakthrough in convergence improvement of the FMM for metallic TM gratings by formulating the so-called Inverse Rule [43]. We recapitulated its proof [1] and remarked that this estimation of pointwise convergence still lacks a physical explanation. Additionally, we showed with the justification of matrix truncation [44] that this convergence is non-uniform. We stated the FMM and its variants [49]: the difference of their respective eigenvalue problems reduces to different representations of the respective Toeplitz matrices of the permittivity profile. In the common FMM variants, the use of the Inverse Rule in 2D is extended with the help of either an automatically generated normal vector field [75] or a Jones polarization basis [2] for 3D problems. However, we did not analyse the FMM including the advanced concept of Adaptive Spatial Resolution [18].

Furthermore, we briefly derived the weak formulation of Maxwell's equations to illustrate the basic ideas of the Finite Element Method [57]. We showed the decomposition into an interior and an exterior problem and their discretizations. Perfectly matched layers were motivated for the solution of the exterior problem and basic statements on the convergence of the FEM led to the illustration of the concept of *hp*-adaptivity [16], where both the mesh size h and the polynomial degree p of the ansatz functions are locally adapted. With the help of a small numerical example we analysed *a priori* p-adaptivity. In addition, we presented a layering algorithm to obtain two-dimensional cross sections needed for FMM simulations from a 3D FEM grid and applied it to an advanced tetrahedral grid.

Before exploring the convergence behaviour of the FMM in various numerical experiments, we verified the unified interface of S^4 and *JCMsuite* for an oblique propagating wave in vacuum and at a simple material interface. Subsequently, we showed that h-adaptivity in the FEM is beneficial for field distributions exhibiting singularities. Moreover, we observed the well-known oscillatoric behaviour of the FMM which can be flattened by smoothing of the permittivity profile [50]. However, our results suggest that this filtering introduces large errors and an extensively increased number of Fourier harmonics is needed to obtain similar results compared to non-smoothed FMM. For the use of a Fast Fourier Transform for the permittivity, we found high error bounds and confirmed findings that closed-form Fourier Transform yields more accurate results.

Since the Gibbs phenomenon is a well-known limit of Fourier Transformation of discontinuous functions, we analysed the solution of the FMM for a binary grating compared to FEM results which model the analytical Fourier Transform of the permittivity in the grating layer but use a polynomial rather than a plane wave basis (as the FMM) for the electromagnetic fields. We showed that the

plane wave basis of the FMM introduces roughly one order of magnitude to the relative errors of the Fourier coefficients. In particular, the error for high diffraction orders increases for both the full FMM and the Fourier Transform material approximation in the FEM. However, we showed that the figure of merit of grating theory and the analysis of Photonic Crystals, the diffraction efficiencies and the reflectance and transmittance spectra respectively, are computed accurately by the FMM. Furthermore, by investigating local near-field error contributions in L^2 norm, we found that errors arise at layer interfaces and are propagated through the whole structure by the scattering matrix formalism of the FMM.

Although the completeness of the plane wave basis set is not rigorously proven for complex permittivity profiles [39], the FMM is often applied to geometries including metallic features. That is why we analysed an EUV mask with metallic scatterers and confirmed findings that errors increase in the metallic regions [34]. Additionally, since the local error contributions of the layer with the metallic features are higher, we proposed using an adaptive number of Fourier harmonics for each layer to improve performance of the FMM. Furthermore, we illustrated convergence improvement of the Inverse Rule for the far-field Fourier coefficients of a dielectric grating. However, we found that the near-field error of electromagnetic field energy reveals larger errors for the Inverse Rule compared to Laurent's Rule which should be further analysed.

We tested performance of staircasing of the FMM [55] for a 2D PhC slab adapted from [3]. We found that the interplay of the additional numerical parameter of the number of layers and the number of Fourier harmonics is generally unpredictable. Furthermore, we showed that the erroneous phase shift increases with the number of layers similar to the findings for the investigated EUV mask. On the other hand, the obtained transmittance spectra of the FMM and the FEM exhibit similar features and their resonance frequency shift can be scaled down.

In the case of 3D simulations, we analysed a checkerboard grating consisting of absorbing material. Performance of the different variants of the FMM depends strongly on the polarization of the illumination. Generally, the naive FMM extension to 3D which dismisses proper Fourier Factorization Rules is slowly convergent. However, the normal vector method and using a Jones vector field both significantly speed up convergence of the FMM, particularly for small numbers of Fourier harmonics in the mid-hundreds (cf. [49]). For this particular example the normal vector method's reflectance and transmittance matches FEM results best.

Nevertheless, for a quadratic pin hole in an absorbing layer, the Jones polarization basis yields much better results than the normal vector method compared to the FEM. Here, we analysed only the 0-th diffraction orders since higher diffraction orders show much slower convergence due to significant field singularities. The relative error (including phase shift) of the Fourier coefficient when using the FMM Jones variant is less than 1% with respect to FEM results. However, we found that subpixel smoothing led to results which differ roughly by a factor of two to FEM and the other FMM formulations. We finished our convergence study of the FMM with a qualitative comparison of the band diagram of a three-dimensional PhC slab simplified from [8]. Again, we found that the Jones polarization yields results which are more accurate than the ones obtained with the normal vector method. Furthermore, we analysed the fulfilment of energy conservation of the different numerical methods and found that for the chosen setup the Jones FMM shows slightly smaller errors and less error distribution than the FEM for comparable computing times.

Excluding the last numerical experiment, we did not present benchmarking results including simulation time and memory consumption. It was the aim of this project to investigate the general applicability of the FMM to a wide range of nano-optical scattering problems. It is known that for sinusoidal gratings, for instance, specialized methods exist [14] which show high performance for a limited class of problems. However, detailed benchmarking requires sophisticated and advanced implementation of both methods which is beyond the scope of this work. Nevertheless, we found that the FMM often yields faster results the quality of which cannot be easily checked self-consistently. On the other hand, the FEM offers more numerical parameters and a well established convergence theory to double-check its findings, yet one has to be more careful with mesh generation, the choice of polynomial degree as well as PML discretization. This flexibility offers the possibility to adjust error tolerances to a specific problem, whereas it is not *a priori* clear which FMM variant is optimally suited.

In conclusion, we found that the FMM yields accurate results for the fields of grating theory when analysis is limited to diffraction efficiencies and the investigation of Photonic Crystals if one is only interested in energy diffraction. However, the near-fields are not as well approximated as when applying the FEM and quality of far-field phase correlations of the Fourier coefficients is also limited. Therefore, the FMM could serve for preliminary results but should be complemented with error-controlled FEM simulations.

Acknowledgements

I would like to thank the following people who supported me throughout my work and made this thesis possible:

... Professor Thomas Judd who enabled me to write my thesis in a collaboration of the University of Tübingen and the Konrad-Zuse-Institute Berlin and who supervised the progress of my studies through numerous e-mails, phone calls, meetings and a visit in Berlin.

... Professor Frank Schmidt who introduced me to the field of nano-optics, supported the organisation of working on my thesis at the Konrad-Zuse-Insitute, let me become a member of his research group, presented many co-workers to me and motivated me and asked the right questions during the time of writing.

... Dr Sven Burger and Mr Martin Hammerschmidt who always answered my countless questions, discussed various numerical and physical problems with me and contributed with their time and knowledge to this work.

... all members and alumni of the nano-optics research group at the Konrad-Zuse-Insitute and JCMwave who assisted me with every problem, provided me insights in their research and introduced me to sophisticated academic code developed during many years.

... Mr Marc Seifried and Mr Markus Kantner who discussed with me for many hours about single-photon sources, electro-optical coupling, coding problems, physics and life in general.

... Ms Lisa Poulikakos who introduced me to the ideas of optical chirality, lead many long and fruitful telephone conversations with me and gave me access to her research.

... Mr Jakob Rosenkrantz de Lasson who enabled me to visit him at his research group, replied extensively to many e-mails about the FMM and kept me up-to-date about his PhD project.

... my whole family and especially my girlfriend, Ms Claudia Horn, who motivated me, listened to me and supported me throughout my studies and the final year of writing this work.

... Ms Kornelia Csink who proofread the manuscript and endured the technical English.

Appendix A

Software Interface

A.1 Software I/O

Table A.1: *Input and output generated by the software packages JCMsuite and S^4. JCMsuite is based on general Cartesian tensor fields and arbitrary discretization of geometry for FEM simulations. S^4 uses layers for geometrical representation and handles x- and y-anistropic materials.*

	JCMsuite[a]	S^4
INPUT	*JCM markup language*	*Lua scripting language*[b]
geometry	discretized grid (triangles, tetrahedrons, bricks, prisms, ...)	z-sorted layers with layer patterns (rectangles, circles, polygons)
materials	rel. permittivity, rel. permeability	rel. permittivity
source (electric plane wave)	3D-Cartesian amplitude and $\boldsymbol{k}$ vector	frequency, s- and p-amplitudes and phases, spherical rotation angles
OUTPUT	*JCM-ASCII / -binary format*	*Lua table*
Fourier Transform	3D-Cartesian Fourier coefficients	(a) z-component of Poynting vector, (b) s- and p-ampltiudes of $\boldsymbol{H}$
$\boldsymbol{E}$-field	export on Cartesian grid	(a) pointwise evaluation, (b) export on Cartesian grid in plane parallel to wave fronts

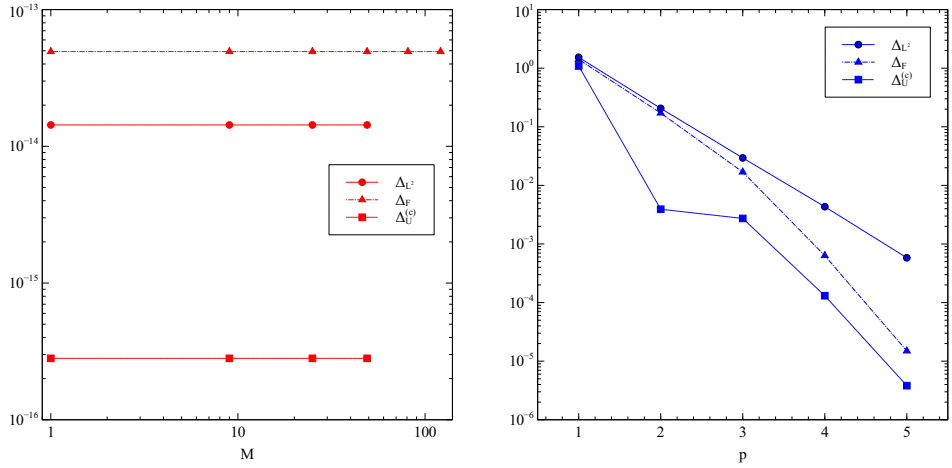
[a] only I/0 of *JCMsuite* relevant for this project is mentioned
[b] S^4 offers recently a python interface as well

Table A.2: *Modified input and output generated by the modified versions of the software packages JCMsuite and S^4. JCMsuite is extended by layering given discretizations (see Sec. III.3). S^4 operates on the same input markup language as JCMsuite. Note that the use of fast $\boldsymbol{E}$ evaluation on a Cartesian grid parallel to the wave fronts in the FMM (see Table A.1) cannot be generalized to conical illumination.*

	JCMsuite	S^4
INPUT		*JCM markup language*
geometry		*JCMsuite crossSections.jcm* obtained by layering algorithm of geometry discretization
materials		*JCMsuite materials.jcm*
source		*JCMsuite sources.jcm* including phase shifts in the z-direction
OUTPUT		
geometry	export cross sections of discretized grid (including zero point shift, automatic cuts at domain interfaces, pointlist of cuts)	
Fourier Transform		3D-Cartesian Fourier coefficients of forward and backward propagating modes
$\boldsymbol{E}$-field		export on Cartesian grid

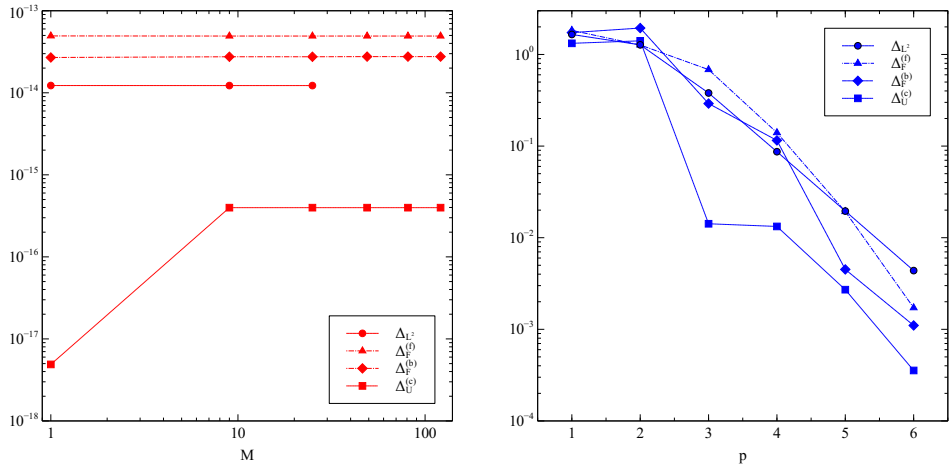
A.2 3D Verification

A.2.1 Analytical Comparison



(a) FMM (b) FEM

Figure A.1: *Convergence of FMM (a) and FEM (b) simulations for a propagating plane wave in 3D. Relative errors of the near-field in L^2 norm Δ_{L^2} (circles), summed relative errors of the Fourier coefficients Δ_F (triangles) and relative errors of the electric field energy $\Delta_U^{(c)}$ (squares) are shown. For their definitions refer to Section II.4. Errors are computed with respect to the analytical values.*



(a) FMM (b) FEM

Figure A.2: *Convergence of FMM (a) and FEM (b) simulations for a propagating plane wave at a material interface in 3D. Relative errors of the near-field in L^2 norm Δ_{L^2} (circles), summed relative errors of the forward propagating Fourier coefficients $\Delta_F^{(f)}$ (triangles) as well as the reflected Fourier coefficients $\Delta_F^{(b)}$ (diamonds) and relative errors of the electric field energy $\Delta_U^{(c)}$ (squares) are shown. Errors are computed with respect to the analytical values. Deviations in the electric field energy ($\Delta_U^{(c)}$) of the FMM for different numbers of harmonics are purely numerical artefacts.*

A.2.2 2D and 3D Comparison

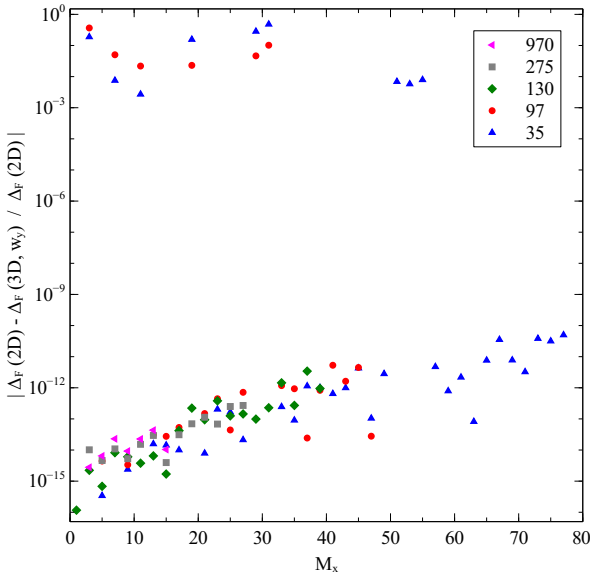


Figure A.3: *Comparison of the interface of S^4 for 2D and 3D simulation of the example in Section IV.2.1.1. This grating is one-dimensional periodic in the x-direction. We use different pitches in the invariant y-direction: $p_y = 970, 275, 130, 97, 35$ nm are plotted with pink left-oriented triangles, grey squares, green diamonds, red circles and blue upward-oriented triangles, respectively. Data is plotted with respect to the number of Fourier harmonics M_x, the k vectors of which are only oriented in the x-direction. In general the total relative Fourier error Δ_F shows that results are equal (up to numerical accuracy 10^{-10}). Nevertheless, for small pitches $p_y = 35, 97$nm particularly for small numbers of Fourier harmonics there are high errors. This is probably caused by an insufficient number of Fourier harmonics in the y-direction which causes these errors.*

Appendix B

Dual Symmetry

We define Dual Symmetry in non-Gaussian view for homogeneous space to be

$$\boldsymbol{E} \to \widetilde{\boldsymbol{E}} = \cos(\theta)\boldsymbol{E} - \sin(\theta)\sqrt{\frac{\mu}{\varepsilon}}\frac{\partial_t^n}{\omega^n}\frac{1}{|\boldsymbol{k}|^m}\boldsymbol{\nabla}^m \times \boldsymbol{H} \tag{B.1}$$

$$=: \cos(\theta)\boldsymbol{E} + \sin(\theta)\boldsymbol{E}_d \tag{B.2}$$

$$\boldsymbol{H} \to \widetilde{\boldsymbol{H}} = \cos(\theta)\boldsymbol{H} + \sin(\theta)\sqrt{\frac{\varepsilon}{\mu}}\frac{\partial_t^n}{\omega^n}\frac{1}{|\boldsymbol{k}|^m}\boldsymbol{\nabla}^m \times \boldsymbol{E} \tag{B.3}$$

$$=: \cos(\theta)\boldsymbol{H} + \sin(\theta)\boldsymbol{H}_d \tag{B.4}$$

where $\theta \in \mathbb{R}$, $m, n \in \mathbb{N}$ and $\boldsymbol{E}_d, \boldsymbol{H}_d$ are the dual electric and the dual magnetic field, respectively. θ is an arbitrary degree of mixture between electric and magnetic fields.

This is a non field theory (cf. [9]) and a non potentials ($\boldsymbol{A}, \phi$ cf. [61]) view on the duality between magnetic and electric fields in electromagnetism.

Dual Symmetry leaves the action $\mathcal{S}$ invariant:

$$\mathcal{S} = \int \boldsymbol{D}\cdot\boldsymbol{E} - \boldsymbol{B}\cdot\boldsymbol{H}\ dx^4 = \int \widetilde{\boldsymbol{D}}\cdot\widetilde{\boldsymbol{E}} - \widetilde{\boldsymbol{B}}\cdot\widetilde{\boldsymbol{H}}\ dx^4 \tag{B.5}$$

Following Noether's Theorem, each Dual Symmetry for $m, n \in \mathbb{N}$ provides a conservation law. Yet only a limited number of conservation laws are directly correlated to basic physical concepts.

The quantities of the conservation law

$$\partial_t \rho_{m,n} + \boldsymbol{\nabla}\cdot\boldsymbol{S}_{m,n} = 0 \tag{B.6}$$

can be obtained with the rule

$$\rho_{m,n} = \widetilde{\boldsymbol{D}}\cdot\widetilde{\boldsymbol{E}} + \widetilde{\boldsymbol{B}}\cdot\widetilde{\boldsymbol{H}} \tag{B.7}$$

$$\boldsymbol{S}_{m,n} = \partial_t\widetilde{\boldsymbol{E}} \times \widetilde{\boldsymbol{H}} \tag{B.8}$$

For $m = 0, n = 0$ we get standard conservation of energy.
For $m = 1, n = 0$ we get standard conservation of linear momentum.
For $m = 0, n = 2$ we get conservation of optical chirality.

Bibliography

[1] B. Anić. *The Fourier-Galerkin Method for Band Structure Computations of 2D and 3D Photonic Crystals*. PhD thesis, Karlsruhe Institute of Technology, 2013.

[2] R. Antos. Fourier Factorization with Complex Polarization Bases in Modeling Optics of Discontinuous Bi-Periodic Structures. *Opt. Express*, 17(9):7269–7274, 2009.

[3] R. Antos and M. Veis. Fourier Factorization with Complex Polarization Bases in the Plane-wave Expansion Method Applied to Two-dimensional Photonic Crystals. *Opt. Express*, 18(26):27511–27524, 2010.

[4] R. Antos, V. Vozda, and M. Veis. Plane Wave Expansion Method Used to Engineer Photonic Crystal Sensors with High Efficiency. *Opt. Express*, 22(3):2562–2577, 2014.

[5] U. Bandelow, H. Gajewski, and R. Hünlich. *Thermodynamics Based Modeling of Edge Emitting Quantum Well Lasers*. Preprint. WIAS, 2004.

[6] P. P. Banerjee and J. M. Jarem. Convergence of Electromagnetic Field Components Across Discontinuous Permittivity Profiles. *J. Opt. Soc. Am. A*, 17(3):491–492, 2000.

[7] G. Bao, L. Cowsar, and W. Masters. *Mathematical Modeling in Optical Science*. Frontiers in Applied Mathematics. Society for Industrial and Applied Mathematics, 2001.

[8] C. Becker, P. Wyss, D. Eisenhauer, J. Probst, V. Preidel, M. Hammerschmidt, and S. Burger. 5×5 cm^2 Silicon Photonic Crystal Slabs on Glass and Plastic Foil Exhibiting Broadband Absorption and High-intensity Near-fields. *Sci. Rep.*, 4, 2014.

[9] K. Y. Bliokh, A. Y. Bekshaev, and F. Nori. Dual Electromagnetism: Helicity, Spin, Momentum and Angular Momentum. *New Journal of Physics*, 15(3):033026, 2013.

[10] K. Y. Bliokh and F. Nori. Characterizing Optical Chirality. *Physical Review A*, 83:021803–1–3, 2011.

[11] M. Bocher. Introduction to the Theory of Fourier's Series. *Annals of Mathematics*, 7(3):81–152, 1906.

[12] T. Bui-Thanh, L. Demkowicz, and O. Ghattas. A Unified Discontinuous Petrov–Galerkin Method and Its Analysis for Friedrichs Systems. *SIAM Journal on Numerical Analysis*, 51(4):1933–1958, 2013.

[13] S. Burger, R. Köhle, and L. Zschiedrich. Benchmark of FEM, Waveguide and FDTD Algorithms for Rigorous Mask Simulation. *SPIE Proceedings*, 5992:599216–1–12, 2005.

[14] J. Chandezon, G. Raoult, and D. Maystre. A New Theoretical Method for Diffraction Gratings and Its Numerical Application. *Journal of Optics*, 11(4):235, 1980.

[15] J. R. de Lasson, P. T. Kristensen, J. Mørk, and N. Gregersen. Roundtrip Matrix Method for Calculating the Leaky Resonant Modes of Open Nanophotonic Structures. *J. Opt. Soc. Am. A*, 31(10):2142–2151, 2014.

[16] L. Demkowicz. *Computing with hp-ADAPTIVE FINITE ELEMENTS: Volume 1. One and Two Dimensional Elliptic and Maxwell Problems*. Chapman & Hall/CRC Applied Mathematics & Nonlinear Science. Taylor & Francis, 2007.

[17] A. Drezet and C. Genet. *Singular and Chiral Nanoplasmonics*, chapter Reciprocity and Optical Chirality. Pan Stanford Publishing, 2014. (in press).

[18] S. Essig and K. Busch. Generation of Adaptive Coordinates and Their Use in the Fourier Modal Method. *Opt. Express*, 18(22):23258–23274, 2010.

[19] J. Fang, B. Liu, Y. Zhao, and X. Zhang. Two-dimensional High Efficiency Thin-film Silicon Solar Cells with a Lateral Light Trapping Architecture. *Sci. Rep.*, 4, 2014.

[20] A. Farjadpour, D. Roundy, A. Rodriguez, M. Ibanescu, P. Bermel, J. D. Joannopoulos, S. G. Johnson, and G. W. Burr. Improving Accuracy by Subpixel Smoothing in the Finite-Difference Time Domain. *Opt. Lett.*, 31(20):2972–2974, 2006.

[21] J. K. Gansel, M. Wegener, S. Burger, and S. Linden. Gold Helix Photonic Metamaterials: A Numerical Parameter Study. *Opt. Express*, 18(2):1059–1069, 2010.

[22] P. Götz, T. Schuster, K. Frenner, S. Rafler, and W. Osten. Normal Vector Method for the RCWA with Automated Vector Field Generation. *Opt. Express*, 16(22):17295–17301, 2008.

[23] G. Granet and B. Guizal. Efficient Implementation of the Coupled-wave Method for Metallic Lamellar Gratings in TM Polarization. *J. Opt. Soc. Am. A*, 13(5):1019–1023, 1996.

[24] N. Gregersen, T. R. Nielsen, B. Tromborg, and J. Mørk. Quality Factors of Nonideal Micro Pillars. *Applied physics letters*, 91(1):011116–011116, 2007.

[25] M. Gschrey, F. Gericke, A. Schüßler, R. Schmidt, J.-H. Schulze, T. Heindel, S. Rodt, A. Strittmatter, and S. Reitzenstein. In Situ Electron-beam Lithography of Deterministic Single-Quantum-Dot Mesa-structures Using Low-temperature Cathodoluminescence Spectroscopy. *Applied Physics Letters*, 102(25), 2013.

[26] R. A. Harris. On the Optical Rotary Dispersion of Polymers. *The Journal of Chemical Physics*, 43(3):959–970, 1965.

[27] J. Hesthaven and T. Warburton. *Nodal Discontinuous Galerkin Methods: Algorithms, Analysis, and Applications*. Texts in Applied Mathematics. Springer, 2008.

[28] V. Heuveline and R. Rannacher. A Posteriori Error Control for Finite Element Approximations of Elliptic Eigenvalue Problems. *Advances in Computational Mathematics*, 15(1-4):107–138, 2001.

[29] E. Hewitt and R. Hewitt. The Gibbs-Wilbraham Phenomenon: An Episode in Fourier Analysis. *Archive for History of Exact Sciences*, 21(2):129–160, 1979.

[30] J. P. Hugonin and P. Lalanne. Perfectly Matched Layers as Nonlinear Coordinate Transforms: a Generalized Formalization. *J. Opt. Soc. Am. A*, 22(9):1844–1849, 2005.

[31] J. D. Jackson. *Classical Electrodynamics*. John Wiley and Sons, 3rd edition, 1998.

[32] JCMsuite. http://www.jcmwave.com, 2014.

[33] H. Kim, J. Park, and B. Lee. *Fourier Modal Method and Its Applications in Computational Nanophotonics*. Taylor & Francis, 2012.

[34] H. Kim, G.-W. Park, and C.-S. Kim. Investigation of the Convergence Behavior with Fluctuation Features in the Fourier ModalAnalysis of a Metallic Grating. *J. Opt. Soc. Korea*, 16(3):196–202, Sep 2012.

[35] S. Kirner, M. Hammerschmidt, C. Schwanke, D. Lockau, S. Calnan, T. Frijnts, S. Neubert, A. Schopke, F. Schmidt, J.-H. Zollondz, A. Heidelberg, B. Stannowski, B. Rech, and R. Schlatmann. Implications of TCO Topography on Intermediate Reflector Design for a-Si/μc-Si Tandem Solar Cells - Experiments and Rigorous Optical Simulations. *Photovoltaics, IEEE Journal of*, 4(1):10–15, 2014.

[36] Konrad-Zuse-Institute Berlin. HTC Cluster. http://typo.zib.de/de/service/it-service/htc.html. Access date: 2014/10/28.

[37] J. Küchenmeister. Three-dimensional Adaptive Coordinate Transformations for the Fourier Modal Method. *Opt. Express*, 22(2):1342–1349, 2014.

[38] P. Lalanne and G. M. Morris. Highly Improved Convergence of the Coupled-wave Method for TM Polarization. *J. Opt. Soc. Am. A*, 13(4):779–784, 1996.

[39] A. Lavrinenko, J. Lægsgaard, N. Gregersen, F. Schmidt, and T. Søndergaard. *Numerical Methods in Photonics*, chapter The Modal Method. Optical Sciences and Applications of Light. Taylor & Francis, 2014.

[40] A. Lavrinenko, J. Lægsgaard, N. Gregersen, F. Schmidt, and T. Søndergaard. *Numerical Methods in Photonics*, chapter Finite Element Method. Optical Sciences and Applications of Light. Taylor & Francis, 2014.

[41] M. Lermer, N. Gregersen, F. Dunzer, S. Reitzenstein, S. Höfling, J. Mørk, L. Worschech, M. Kamp, and A. Forchel. Bloch-Wave Engineering of Quantum Dot Micropillars for Cavity Quantum Electrodynamics Experiments. *Phys. Rev. Lett.*, 108:057402, 2012.

[42] L. Li. Formulation and Comparison of Two Recursive Matrix Algorithms for Modeling Layered Diffraction Gratings. *J. Opt. Soc. Am. A*, 13(5):1024–1035, 1996.

[43] L. Li. Use of Fourier Series in the Analysis of Discontinuous Periodic Structures. *J. Opt. Soc. Am. A*, 13(9):1870–1876, 1996.

[44] L. Li. Justification of Matrix Truncation in the Modal Methods of Diffraction Gratings. *Journal of Optics A: Pure and Applied Optics*, 1(4):531, 1999.

[45] L. Li. Convergence of Electromagnetic Field Components Across Discontinuous Permittivity Profiles: Comment. *J. Opt. Soc. Am. A*, 19(7):1443–1444, 2002.

[46] L. Li and C. W. Haggans. Convergence of the Coupled-wave Method for Metallic Lamellar Diffraction Gratings. *J. Opt. Soc. Am. A*, 10(6):1184–1189, 1993.

[47] D. M. Lipkin. Existence of a New Conservation Law in Electromagnetic Theory. *Journal of Mathematical Physics*, 5:696, 1964.

[48] S. Liu, Y. Ma, X. Chen, and C. Zhang. Estimation of the Convergence Order of Rigorous Coupled-wave Analysis for Binary Gratings in Optical Critical Dimension Metrology. *Optical Engineering*, 51(8):081504–1–081504–7, 2012.

[49] V. Liu. *Computational Electromagnetics for Nanophotonic Design and Discovery*. PhD thesis, Stanford University, 2014.

[50] V. Liu and S. Fan. S^4 : A Free Electromagnetic Solver for Layered Periodic Structures. *Computer Physics Communications*, 183(10):2233 – 2244, 2012.

[51] D. Maystre. Theory of Woods Anomalies. In *Plasmonics*, pages 39–83. Springer, 2012.

[52] K. McInturff and P. S. Simon. The Fourier Transform of Linearly Varying Functions with Polygonal Support. *IEEE transactions on antennas and propagation*, 39:1441–1443, 1991.

[53] K. M. McPeak, C. D. van Engers, M. Blome, J. H. Park, S. Burger, M. A. Goslvez, A. Faridi, Y. R. Ries, A. Sahu, and D. J. Norris. Complex Chiral Colloids and Surfaces via High-Index Off-Cut Silicon. *Nano Letters*, 14(5):2934–2940, 2014.

[54] M. G. Moharam and T. K. Gaylord. Rigorous Coupled-wave Analysis of Planar-grating Diffraction. *J. Opt. Soc. Am.*, 71(7):811–818, 1981.

[55] M. G. Moharam and T. K. Gaylord. Diffraction Analysis of Dielectric Surface-relief Gratings. *J. Opt. Soc. Am.*, 72(10):1385–1392, 1982.

[56] M. Moharam, E. B. Grann, and D. A. Pommet. Formulation for Stable and Efficient Implementation of the Rigorous Coupled-wave Analysis of Binary Gratings. *J. Opt. Soc. Am. A*, 12:1068–1076, 1995.

[57] P. Monk. *Finite Element Methods for Maxwell's Equations*. Numerical Mathematics and Scientific Computation. Clarendon Press, 2003.

[58] M. Neviere and E. Popov. *Light Propagation in Periodic Media: Differential Theory and Design*. Optical Science and Engineering. Taylor & Francis, 2002.

[59] L. Novotny and B. Hecht. *Principles of Nano-Optics*. Principles of Nano-optics. Cambridge University Press, 2006.

[60] E. D. Palik. *Handbook of Optical Constants of Solids*, volume 3. Academic Press, 1998.

[61] T. G. Philbin. Erratum: Electromagnetic Energy Momentum in Dispersive Media. *Phys. Rev. A*, 85:059902, 2012.

[62] T. G. Philbin. Lipkin's Conservation Law, Noether's Theorem, and the Relation to Optical Helicity. *Physical Review A*, 87(4):043843, 2013.

[63] E. Popov and M. Nevière. Grating Theory: New Equations in Fourier Space Leading to Fast Converging Results for TM Polarization. *J. Opt. Soc. Am. A*, 17(10):1773–1784, 2000.

[64] E. Popov, M. Nevière, B. Gralak, and G. Tayeb. Staircase Approximation Validity for Arbitrary-shaped Gratings. *J. Opt. Soc. Am. A*, 19(1):33–42, Jan 2002.

[65] W. Press. *Numerical Recipes in C: The Art of Scientific Computing*. Number 4. Cambridge University Press, 1992.

[66] A. Quarteroni, R. Sacco, and F. Saleri. *Numerical Mathematics*. Texts in Applied Mathematics. Springer, 2007.

[67] S. Ragusa and M. Baylin. Electromagnetic First-Order Conservation Laws in a Medium. *Il Nuovo Cimento B*, 1991.

[68] L. Rayleigh. On the Dynamical Theory of Gratings. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 79(532):399–416, 1907.

[69] D. Rittenhouse. An Optical Problem Proposed by F. Hopkinson and Solved. *J. Am. Phil. Soc*, 201:202–206, 1786.

[70] S^4 (Stanford Stratified Structure Solver). http://web.stanford.edu/group/fan/S4, 2014.

[71] H. Sagan. *Boundary and Eigenvalue Problems in Mathematical Physics*. Courier Dover Publications, 2012.

[72] F. P. Sayer. The Eigenvalue Problem for Infinite Systems of Linear Equations. *Mathematical Proceedings of the Cambridge Philosophical Society*, 82:269–273, 9 1977.

[73] M. Schäferling, D. Dregely, M. Hentschel, and H. Giessen. Tailoring Enhanced Optical Chirality: Design Principles for Chiral Plasmonic Nanostructures. *Phys. Rev. X*, 2:031010, 2012.

[74] M. Schäferling, X. Yin, N. Engheta, and H. Giessen. Helical Plasmonic Nanostructures as Prototypical Chiral Near-Field Sources. *ACS Photonics*, 1(6):530–537, 2014.

[75] T. Schuster, J. Ruoff, N. Kerwien, S. Rafler, and W. Osten. Normal Vector Method for Convergence Improvement Using the RCWA for Crossed Gratings. *J. Opt. Soc. Am. A*, 24(9):2880–2890, 2007.

[76] M. Seifried. Entwicklung Determistischer Quantenpunkt-Mikrolinsen mit erhöhter Auskopplungs-Effizienz. Master's thesis, TU Berlin, 2014.

[77] Y. Shen, D. Ye, I. Celanovic, S. G. Johnson, J. D. Joannopoulos, and M. Soljai. Optical Broadband Angular Selectivity. *Science*, 343(6178):1499–1501, 2014.

[78] H. S. Sözüer, J. W. Haus, and R. Inguva. Photonic Bands: Convergence Problems with the Plane-wave Method. *Phys. Rev. B*, 45:13962–13972, 1992.

[79] T. Tamir, H. C. Wang, and A. A. Oliner. Wave Propagation in Sinusoidally Stratified Dielectric Media. *IEEE Trans. Mirow. Theory Technol.*, 12:323–35, 1964.

[80] Y. Tang and A. E. Cohen. Optical Chirality and Its Interaction with Matter. *Phys. Rev. Lett.*, 104:163901, 2010.

[81] F. Wang. *Development of a Fast Converging Hybrid Method for Analyzing Three-dimensional Doubly Period Structures*. PhD thesis, The Ohio State University, 2013.

[82] WIAS-TeSCA. https://www.wias-berlin.de/software/tesca, 2014.

[83] H. Widom. *Studies in Real and Complex Analysis*, volume 3 of *Studies in Mathematics*, chapter Toeplitz matrices, pages 179–212. Mathematical Association of America, 1965.

[84] B. Wohlfeil, S. Burger, C. Stamatiadis, J. Pomplun, F. Schmidt, L. Zimmermann, and K. Petermann. Numerical Simulation of Grating Couplers for Mode Multiplexed Systems. In *SPIE OPTO*, pages 89880K–89880K. International Society for Optics and Photonics, 2014.

[85] R. Wood. XLII. On a Remarkable Case of Uneven Distribution of Light in a Diffraction Grating Spectrum. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 4(21):396–402, 1902.

[86] C. Yeh, K. F. Casey, and Z. A. Kaprielian. Transverse Magnetic Wave Propagation in Sinusoidally Stratified Dielectric Media. *IEEE Trans. Mirow. Theory Technol.*, 13:297–302, 1965.

[87] L. Zschiedrich, S. Burger, B. Kettner, and F. Schmidt. Advanced Finite Element Method for Nano-resonators. In *Physics and Simulation of Optoelectronic Devices XIV*, volume 6115 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pages 164–174, 2006.

[88] L. Zschiedrich, S. Burger, A. Schadle, and F. Schmidt. Domain Decomposition Method for Electromagnetic Scattering Problems: Application to EUV Lithography. In *Numerical Simulation of Optoelectronic Devices, 2005. NUSOD '05. Proceedings of the 5th International Conference on*, pages 55–56, 2005.

[89] L. W. Zschiedrich. *Transparent Boundary Conditions for Maxwell's Equations*. PhD thesis, Freie Universität Berlin, 2009.

[90] D. Zwillinger. *CRC Standard Mathematical Tables and Formulae, 31st Edition*. Discrete Mathematics and Its Applications. Taylor & Francis, 2002.

[91] A. Zygmund. *Trigonometric Series*. Number Bd. 1 in Cambridge Mathematical Library. Cambridge University Press, 2002.