Theoretical and computational insights into the nonlinear optics of nanostructured bulk and 2D materials

Martin Weismann

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
of
University College London.

Department of Electronic and Electrical Engineering
University College London

June 1, 2016

I, Martin Weismann, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

In this thesis, a comprehensive analytical and computational study of linear and nonlinear optical response of nanostructured two-dimensional (2D) and bulk materials is presented. The new numerical methods developed in this thesis facilitate an efficient and accurate design of new artificial optical materials and novel nonlinear optical devices. Moreover, the presented results and conclusions can provide a deeper theoretical understanding of different resonant, nonlinear optical phenomena in photonic nanostructures.

Two computational electromagnetic methods to calculate the interaction of light with linear and nonlinear diffraction gratings and more general periodic structures have been developed. An efficient formulation of the rigorous coupled-wave analysis (RCWA), a modal frequency domain method, for accurate near-field calculations and for complex oblique structures has been proposed. This method has been implemented into a powerful commercial software and applied to calculate diffraction in several nanophotonic devices highly relevant to practical applications. Beyond this commercial implementation, the RCWA has been extended to describe nonlinear optical effects in nanostructured graphene and transition metal dichalcogenide monolayers, with second- and third-harmonic generation being the most important nonlinear processes. A key feature of this formulation is that it is independent of the height of the 2D material and only requires knowledge of its linear and nonlinear optical properties. Using this method, plasmon resonances of nanostructured graphene have been investigated and tuneable Fano resonances have been explored to increase the efficiency of nonlinear optical effects in photonic heterostructures comprising transition metal dichalcogenide monolayers and nanopatterned slab waveguides.

The second thrust of the thesis was devoted to extending the so-called generalised source method (GSM) to the area of nonlinear optics. In particular, its mathematical formulation has been extended to incorporate second- and third-order nonlinear optical effects. The proposed nonlinear GSM has been used to design and optimise multi-resonant photonic devices made of nonlinear optical materials. In addition, this advanced computational method facilitated the understanding of strong nonlinear optical activity in plasmonic nanostructures and explained the multipolar nonlinear optical response of certain nonlinear metasurfaces.

Acknowledgements

I would like to express my absolute gratitude to my supervisor, Dr. Nicolae C. Panoiu. He encouraged me to aim for the highest standards in every project I tackled throughout my PhD and helped me to achieve them through his competent counsel, his tremendous expertise and shear unlimited patience. I consider myself fortunate to have experienced his guidance and I will always value our professional relationship.

My extended gratitude goes to my second supervisor, Dr. F. Aníbal Fernández, for the advice he offered to me throughout my PhD and for his constructive criticism of my reports and my work in general. I also want to show gratitude towards Dr. Dominic F. G. Gallagher, who offered me the opportunity to implement my research in a commercial environment at Photon Design, and to experience the professional world outside of academia.

My sincere thanks go to my close collaborators at UCL, Prof. Paul A. Warburton and Nuno V. S. Braz. Our regular meetings gave me valuable insights into the world of fabrication and experimental techniques, and helped me to connect my theoretical and computational work with the real world. In a similar fashion, I want thank all my collaborators and co-authors, in particular Prof. Yuri S. Kivshar, Prof. Dragomir N. Neshev and Dr. Sergey Kruk from the Australian National University for their hospitality during my academic visit of the Nonlinear Physics Centre in Canberra. I also want to express my gratitude to Prof. Shuang Zhang and Dr. Guixin Li from the University of Birmingham for our successful collaboration. It was a privilege and pleasure working with you all.

I thank all members of Dr. Panoiu's group, in particular Spyros Lavdas, Jie You, Dan Timbrell and Abiola Oladipo, who have accompanied me throughout my PhD and made the working days (and nights) in 708 so much more enjoyable. I am also glad about my amazing colleagues at Photon Design and want to thank Vincent Brulis and Richard Lycett for making my time in Oxford unforgettable.

Finally, I want thank my family for their unconditional support throughout all my life.

I appreciate the financial support through a UCL Impact Award studentship, funded by UCL and Photon Design, through the European Research Council and the Royal Society, and I acknowledge the use of the UCL Legion High Performance Computing Facility (Legion@UCL), and associated support services, in the completion of this work.

Articles in peer-reviewed journals

1. **M. Weismann** and N. C. Panoiu, "Theoretical and computational analysis of second- and third-harmonic generation in periodically patterned 2D-3D heteromaterials," in review by Phys. Rev. B.

2. S. Chen, F. Zeuner, **M. Weismann**, B. Reineke, G. Li, V. K. Valev, K. W. Cheah, N. C. Panoiu, T. Zentgraf, and S. Zhang, "Giant Nonlinear Optical Activity of Achiral Origin in Planar Metasurfaces with Quadratic and Cubic Nonlinearities," Adv. Mater. **28**, 29922999 (2016).

3. **M. Weismann**, D. F. G. Gallagher, and N. C. Panoiu, "Accurate near-field calculation in the rigorous coupled-wave analysis method," J. Opt. **17**, 125612 (2015).

4. S. Kruk, **M. Weismann**, A. Y. Bykov, E. A. Mamonov, I. A. Kolmychek, T. Murzina, N. C. Panoiu, D. N. Neshev, and Yuri S. Kivshar, "Enhanced Magnetic Second-Harmonic Generation from Resonant Metasurfaces," ACS Photonics **2**, 1007–1012 (2015).

5. **M. Weismann**, D. F. G. Gallagher, and N. C. Panoiu, "Nonlinear generalized source method for modeling second-harmonic generation in diffraction gratings," J. Opt. Soc. Am. B **32**, 523–533 (2015).

6. **M. Weismann**, D. F. G. Gallagher, and N. C. Panoiu, "Generalized source method for modeling nonlinear diffraction in planar periodic structures," Proc. SPIE **9131**, 913115-11 (2014).

Contributions to peer-reviewed conferences

1. **M. Weismann** and N. C. Panoiu, "Computational Study of Second- and Third-Harmonic Generation in Periodically Patterned 2D-3D Heteromaterials," submitted to IEEE NANO 2016 International Conference on Nanotechnology, Sendai, Japan, 2016.

2. N. C. Panoiu and **M. Weismann**, "Enhanced Quadratic and Cubic Optical Nonlinearities in Photonic Structures Containing Nanopatterned Graphene and other 2D Materials," submitted to Metamaterials' 2016 10th International Congress on Advanced Electromagnetic Materials in Microwaves and Optics, Crete, Greece, 2016.

3. **M. Weismann** and N. C. Panoiu, "Comparative study of resonantly enhanced optical harmonic generation in nanostructured 2D materials," SPIE Optics and Photonics, San Diego, USA, 2016.

4. N. C. Panoiu and **M. Weismann**, "Computational Modeling of Nonlinear Optical Response of 2D-3D Heteromaterials" [invited], ICEAA 2016 International Conference on Electromagnetics in Advanced Applications, Cairns, Australia, 2016.

5. G. Li, S. Chen, F. Zeuner, **M. Weismann**, B. Reineke, V. K. Valev, K. W. Cheah, N. C. Panoiu, S. Zhang, and T. Zentgraf, "Giant Nonlinear Optical Activity from Planar Metasurfaces" [invited], Metamaterials' 2016 10th International Congress on Advanced Electromagnetic Materials in Microwaves and Optics, Crete, Greece, 2016.

6. N. C. Panoiu and **M. Weismann**, "Computational modeling of higher-harmonic generation in periodic 2D-3D heteromaterials" [invited], PIERS 2016, Shanghai, China, 2016.

7. N. C. Panoiu and **M. Weismann**, "Enhanced optical nonlinearities in periodic photonic nanostructures containing patterned graphene and other 2D materials" [invited], META'16 7th International Conference on Metamaterials, Photonic Crystals and Plasmonics, Malaga, Spain, 2016.

8. **M. Weismann** and N. C. Panoiu, "Numerical modelling of structured nonlinear 2D materials," The Frontiers of Materials Modelling - Thomas Young Centre 10th Anniversary Symposium, London, England, 2016

9. **M. Weismann** and N. C. Panoiu, "Computational method for modelling nonlinear optical response of periodically patterned 2D materials," SPIE Photonics Europe, Brussels, Belgium, 2016.

10. S. Kruk, **M. Weismann**, A. Y. Bykov, E. A. Mamonov, I. A. Kolmychek, T. Murzina, N. C. Panoiu, D. N. Neshev, and Y. S. Kivshar, "Enhanced Magnetic Second-Harmonic Generation from Resonant Metasurfaces," CLEO 2015, San Jose, USA, 2015

11. **M. Weismann**, D. F. G. Gallagher, and N. C. Panoiu, "A comparative study of second-harmonic generation in plasmonic and dielectric gratings made of centrosymmetric materials," SPIE Optics and Photonics, San Diego, USA, 2015.

12. **M. Weismann**, D. F. G. Gallagher, and N. C. Panoiu, "Computational study of second-harmonic generation in diffraction gratings made of centrosymmetric materials using the nonlinear generalized source method," META'15 6th International Conference on Metamaterials, Photonic Crystals and Plasmonics, New York, USA, 2015.

13. **M. Weismann**, D. F. G. Gallagher, and N. C. Panoiu, "Nonlinear generalized source method for modeling bulk and surface second harmonic generation in diffraction gratings," Photon 14, London, UK, 2015.

14. D. Timbrell, **M. Weismann**, N. V. S. Braz, P. A. Warburton, and N. C. Panoiu, "Toroidal dipoles generated in nonlinear plasmonic metamaterials," Metamaterials' 2015 9th International Congress on Advanced Electromagnetic Materials in Microwaves and Optics, Oxford, UK, 2015.

15. D. Timbrell, N. V. S. Braz, **M. Weismann**, P. A. Warburton, and N. C. Panoiu, "Resonantly Enhanced Nonlinear Toroidal Moments in Plasmonic Metamaterials," PIERS 2015, Prague, Czech Republic, 2015.

16. I. A. Kolmychek, E. A. Mamonov, A. Y. Bykov, T. Murzina, S. Kruk, D. N. Neshev, **M. Weismann**, N. C. Panoiu, and Y. S. Kivshar, "Nonlinear-Optical Studies of Magnetic Dipole Metamaterials," Frontiers in Optics 2014, Tuscon, USA, 2014.

17. T. Murzina, S. Kruk. A. Y. Bykov, I. A. Kolmychek. E. A. Mamonov, **M. Weismann**, and N. C. Panoiu, "Nonlinear Spectroscopy of Magnetic Dipole Metasurfaces," Metamaterials' 2014 8th International Congress on Advanced Electromagnetic Materials in Microwaves and Optics, Copenhagen, Denmark, 2014.

18. **M. Weismann** and N. C. Panoiu, "Fourier modal method for efficient modeling of periodically structured metasurfaces with unit cells of arbitrary shape," Metamaterials' 2014 8th International Congress on Advanced Electromagnetic Materials in Microwaves and Optics, Copenhagen, Denmark, 2014.

19. **M. Weismann**, D. F. G. Gallagher, and N. C. Panoiu, "Nonlinear generalized source method for modeling second-harmonic generation in 2D diffraction gratings," SPIE Photonics Europe, Brussels, Belgium, 2014.

Other presentations

1. **M. Weismann**, D. F. G. Gallagher, and N. C. Panoiu, Computational modelling of linear and nonlinear diffraction gratings – overview and case studies, Barlow Memorial Lecture 2015 and Research Poster Presentation, UCL, London, 2015

2. **M. Weismann**, D. F. G. Gallagher, and N. C. Panoiu, Application of a non-linear generalized source method to the design of phase-matched, doubly resonant nonlinear GaAs Bragg waveguides, Mildner Memorial Lecture 2014 and Research Poster Presentation, UCL, London, 2014

3. **M. Weismann**, D. F. G. Gallagher, and N. C. Panoiu, Computational study of diffraction gratings based on the rigorous coupled-wave analysis method, Barlow Memorial Lecture 2013 and Research Poster Presentation, UCL, London, 2013

List of notations and acronyms

1D, 2D, 3D	one-dimensional, two-dimensional, three-dimensional
2DM	two-dimensional material
CAD	computer-aided design
CD	circular dichroism
EM	electromagnetic
ϵ_0	electric permittivity of vacuum ($8.854 \times 10^{-12}\,\mathrm{F\,m^{-1}}$)
ϵ	electric permittivity
ϵ_r	relative electric permittivity (ϵ/ϵ_0)
FF	fundamental frequency
FDTD	finite-difference time-domain method
FETD	finite-element time-domain
FFT	fast Fourier transform
FWM	four-wave mixing
GSM	generalised source method
λ	wavelength
Λ	grating period
ME	Maxwell's equations
μ_0	magnetic permeability of vacuum ($1.257 \times 10^{-6}\,\mathrm{H\,m^{-1}}$)
μ	magnetic permeability
μ_r	relative magnetic permeability (μ/μ_0)
NVF	normal vector field
ODE	ordinary differential equation
RCWA	rigorous coupled-wave analysis
SHG, SH	second-harmonic generation, second-harmonic
TE, TM	transverse electric, transverse magnetic
THG, TH	third-harmonic generation, third-harmonic
TMDC	transition metal dichalcogenide
ω	angular frequency
$\chi^{(1)}$	linear susceptibility
$\chi^{(2)}$	second-order susceptibility
$\chi^{(3)}$	third-order susceptibility

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Photonics, the science of light and its applications, is one of the most dynamic areas of current research in physics and engineering. Electromagnetic and photonic principles are deeply embedded in modern telecommunication systems, medical devices, consumer electronics, renewable energies and lighting systems. In recognition of the fundamental importance of light to our modern way of life, last year has been celebrated by the United Nations as the International Year of Light and Light-based Technologies. This recognition of the importance of light and its impact on our life, however, does not mark the peak of research in photonics. Instead, the potential economic and social impact of future applications and technologies based on photonics, such as optical cloaking, nano-scale lasing or integrated silicon photonics, can hardly be estimated, so that photonics is expected to remain one of the key technologies of the 21^{st} century playing the central role which electronics played in the last century.

A large class of these light-based technologies, such as optical sensing, display technologies, non-invasive medical diagnosis, and fibre-optical communication, depend on nonlinear optical phenomena, such as higher-harmonic generation, in microscopic structures. While the theory of linear interaction of light with matter is well understood, the investigation of these nonlinear optical processes in nanostructured materials is far more challenging and less understood, due to the higher complexity of the underlying mathematical models and the physical intricacies pertaining to how they manifest in the bulk or at the surface of nonlinear media. Experimental investigations are equally problematic, because nonlinear optical effects are extremely difficult to measure let alone harness due to their low intensity and their high sensitivity to geometric imperfections or defects in nanostructures. Despite the fact that our capability to fabricate optical structures at the micro- and nanometer scale has improved greatly throughout the last decades and complex nonlinear optical devices can be fabricated with ever higher accuracy, advanced nanofabrication is often difficult to perform, time-consuming, and expensive.

To overcome these challenges, highly sophisticated computational methods comple-

ment the experimental investigation of optical structures and devices, and aid us in the concrete understanding of theoretical models or experimental results. Computational investigation offers a deeper insight into the electromagnetic behaviour of photonic nanostructures than most experimental techniques, and numerical modelling even drives the development of novel applications. It can accelerate their design process and make it more efficient and affordable by replacing the expensive cycle of fabrication, test, and redesign in early stages of research and development. This is currently exemplified by the ongoing research on optical metamaterials, artificially designed nanostructures with tailored optical properties, whose design greatly benefits from the capabilities of powerful computational software tools to simulate their intricate interaction with light. In this fashion, numerical simulation has been an instrumental tool in electromagnetic research for over half a century, and continues to drive innovation and progress in photonics. However, similar to the way nanofabrication has to improve with the development of novel concepts, materials, applications and devices, also the capability of methods in computational electromagnetics has to progress.

A concrete example for this demand for advanced numerical methods is given by the advent of two-dimensional materials in the last decade, most notably graphene, a monolayer of carbon atoms arranged in a hexagonal lattice. Two-dimensional materials have remarkable optical and optoelectronic properties, and they offer huge potential to enhance or miniaturise the functionality of conventional bulk materials, or enable completely new functionality at the nanoscale. Novel fabrication techniques allow the preparation of complex devices incorporating nanostructured 2D materials, which further increases the need for their accurate optical simulation. In particular, efficient and rigorous computational modelling of nonlinear optical effects in devices containing 2D materials has been largely not been addressed. It is thus a main objective of this thesis to develop numerical methods for the investigation of nonlinear optical phenomena in nanostructures containing bulk and 2D materials. The next section will define the objectives of this thesis in more detail.

1.1 Main objectives of this thesis

Two main objectives have been formulated for this thesis. The first objective is the development of efficient and accurate computational methods for modelling periodic optical structures containing nonlinear optical materials based on a solid theoretical description of nonlinear optics. Importantly, these numerical methods shall be capable of simulating different, dielectric and metallic materials, both three-dimensional (3D) and two-dimensional in nature. The second objective has been the application of these methods to develop efficient nonlinear nanostructures, and assist the experimental investigation of linear and nonlinear optical phenomena in photonic devices, in particular in nonlinear metasurfaces, by numerical simulations. These two objectives will be specified more precisely now.

The prerequisite for the development of a numerical method is a solid understanding of the underlying physics. To meet this challenge, I will present mathematical models for linear and nonlinear wave-interaction with bulk and 2D materials and give a broad definition for the diffraction grating, which is the generic structure to be investigated. Based on this scope of problems, two numerical methods, the rigorous coupled wave-analysis (RCWA) and the generalised source method (GSM), which are both based on Fourier series expansion, will be selected as the general basis for all numerical investigations in this thesis. The mathematical theory of trigonometric series, which is fundamental to the accuracy and convergence of the RCWA and the GSM, will be revisited and put into the context of these two methods.

Armed with this physical and mathematical knowledge, I will derive a modular and flexible formulation of the RCWA for the linear interaction of light with matter, which incorporates recent and relevant improvements to this numerical method, which are beneficial to the goal of modelling nonlinear phenomena with RCWA. In particular, I will put special focus on the accuracy of the calculated optical near-field, as it is of central importance for nonlinear optical processes, and show the benefits of the proposed accurate near-field evaluation in RCWA. I will explain the details of the computer implementation of this advanced formulation of the RCWA into a unique commercial software tool, `OmniSim RCWA`, and discuss its computational capabilities and algorithmic details. Beyond this commercial implementation, I will derive, implement and benchmark a modification of the RCWA which allows the rigorous numerical simulation of arbitrary, structured nonlinear 2D materials, which are embedded in larger periodic devices. Using the nonlinear RCWA for 2D materials, I will design periodically patterned nanostructures comprising 2D materials, including graphene and transition metal dichalcogenide monolayer materials, to enhance their nonlinear efficiency, after providing a comprehensive overview of their linear and nonlinear optical properties.

In order to numerically model the optical nonlinearities in conventional bulk materials, I derive a nonlinear modification of the GSM and focus on modelling second-harmonic generation and third-harmonic generation in periodically nanostructured materials.Subsequently, I will apply this novel numerical method to model the nonlinear optical activity in 2D-periodic, plasmonic nanostructures to complement their experimental investigation. This comparison will not only provide validation of the numerical method and its implementation but also give deep insight into the achiral origin of the nonlinear dichroism in these nanostructures. Additionally, I will analyse the resonant enhancement of second-harmonic generation in a resonant metasurface. To this end, I will derive an analytical model for multipolar radiation in periodic structures, which is used to investigate the physical origin of this enhanced nonlinearity by comparison to numerical results from rigorous

RCWA and GSM simulations.

1.2 Outline

Chapter 2 presents literature, physics background, reviews important properties of diffraction gratings and general periodic structures and nonlinear optical processes, introduces 2D materials and their optical properties and possible applications. General numerical methods of computational electromagnetics will be reviewed and the choice of two specialised numerical methods, the rigorous coupled-wave analysis and the generalised source method, will be motivated. This chapter serves as a common denominator for the later derivations and investigations, and the following chapters will pick up specific concepts from this chapter, elaborate and relate back to them.

Chapter 3 rigorously derives the RCWA, the first numerical method developed in this work. An overview of the most important developments of RCWA is presented, and the mathematical background for Fourier series, which is the basic discretisation in the RCWA and GSM, will be given to facilitate this derivation. Moreover, this chapter formulates and solves the Fourier factorisation problem, and shows its importance for convergence of both the RCWA and the GSM. The derivation of the RCWA also includes the numerically stable scattering matrix formulation for complex, metallic or dielectric, multilayered periodic structures and first numerical examples illustrating the algorithmic choices made in this formulation and their benefits to convergence and accuracy of the RCWA.

The accuracy with which the optical near-field can be calculated by the RCWA is investigated in Chapter 4, and it explains the observed slow convergence and numerical artefacts from which it suffers, namely unphysical oscillations at material boundaries due to the Gibbs phenomenon. In order to alleviate these shortcomings, a mathematical formulation for accurate near-field calculation in RCWA, for one- and two-dimensional straight and slanted diffraction gratings is introduced. This accurate near-field computational approach is tested and evaluated for several representative test-structures and configurations in order to illustrate the advantages provided by the proposed modified formulation of the RCWA.

Chapter 5 describes the computer implementation of the RCWA in the advanced formulation of Chapters 3 and 4 into a sophisticated, commercial software tool, OmniSim RCWA. This powerful and user friendly software allows the accurate and efficient modelling of arbitrary diffraction grating structures and optical metamaterials. The logical program flow of OmniSim RCWA will be presented and special attention will be devoted to specific algorithmic choices for normal vector field generation and scattering matrix calculations.

The focus of the remainder to this thesis, which goes beyond this commercially available work, is the numerical modelling of optical nonlinearities in photonic structures arising in conventional bulk materials and novel 2D materials. Thus, Chapter 6 derives the RCWA

for linear and nonlinear optical 2D materials in complex periodic structures. This is complemented by a comprehensive description of the optical properties of graphene and transition metal dichalcogenide monolayer materials, their symmetry properties and the resulting nonlinearities, namely third- and second-harmonic generation, respectively. The proposed nonlinear RCWA is used to investigate multiresonant devices for enhanced and tunable higher-harmonic generation from these 2D materials, and the interplay of plasmonic and geometric resonances in a combined graphene-slab waveguide structure is analysed.

A different numerical method, the nonlinear GSM, is derived in Chapter 7 to model optical nonlinearities in structured 3D materials constituting diffraction gratings. It will be shown how the nonlinear GSM algorithmically incorporates the quadratic nonlinearity of non-centrosymmetric materials and its convergence will be investigated in a set of benchmark calculations. Thereafter, it is employed to demonstrate strong enhancement of SHG one- and two-dimensional optical gratings resonantly coupled to a slab waveguide.

Chapter 8 applies the nonlinear GSM to model THG by a plasmonic Gammadion-type nanostructure and describes a theoretical model for the calculation of the effective nonlinear susceptibility of nonlinear Gammadion- and Trisceli-type metasurfaces with three- and four-fold rotational symmetries, respectively. Simulations and experimental observations are shown to agree well and confirm that a giant nonlinear optical activity in form of near-unity nonlinear circular dichroism can be achieved in these structures.

Chapter 9 studies the second-order nonlinear response from resonant metasurfaces composed of metallo-dielectric nanodisks. In addition to its experimental and numerical verification, a theory of multipolar radiation from periodic structures is developed and used to explain the origin of the resonant enhancement of nonlinear harmonic generation in the metal-dielectric meta atoms.

The main conclusions of this thesis will be drawn in Chapter 10 and the contributions of this work to the fields of computational electromagnetics and nonlinear optics will be highlighted. Future prospects and ideas for extending this work are also presented in this concluding chapter.

Chapter 2

Background

2.1 Introduction

The functionality and versatility of modern photonic devices arise from an in-depth theoretical understanding of photonic principles and the the availability of state-of-the-art nanofabrication techniques. If these two prerequisites are given, then a photonic device can fulfil its specific role as part of a larger photonic system. The underlying physics of most photonic devices are described by Maxwell's equations, a set of complex mathematical partial differential equations. They lack analytical solutions in most practical situations and become even more complex when nonlinear optical effects are considered. These challenges can be met by employing sophisticated computational methods and their software implementation as these tools enable one to achieve an accurate description of the dynamics of the electromagnetic field and its interaction with matter.

This chapter describes the theoretical background upon which this thesis builds its original work. It serves as a common denominator for the remaining parts of the thesis, which will subsequently refer back to the concepts introduced here and further elaborate on them. In particular, Section 2.2 will discuss Maxwell's equations, the mathematical formalism that governs electromagnetism, and explain their most important properties. Section 2.3 introduces the nomenclature and theoretical aspects of diffraction gratings, a particular class of periodic structures, which are the main device application of the computational methods developed in this thesis. Subsequently, a short introduction to the theory of nonlinear optics will be given in Section 2.4. The most widely used numerical methods of computational electromagnetics will be classified and characterised in Section 2.5, where a particular focus is devoted to computational methods for modelling diffraction gratings. Thereafter, Section 2.6 will illustrate a wide range of important applications of diffraction gratings and nonlinear optical phenomena. Finally, Section 2.7 will provide a short introduction to 2D materials and their potential applications in photonics.

2.2 Maxwell's equations

The physical principles of optics and electromagnetic radiation are described by Maxwell's equations, which were initially formulated by James Clerk Maxwell in Ref. [1]. The macroscopic, time dependent formulation of Maxwell's equations (ME) in differential form reads [2]

$$\nabla \cdot \boldsymbol{\mathscr{D}}(\mathbf{r},t) = \rho(\mathbf{r},t), \tag{2.1a}$$

$$\nabla \cdot \boldsymbol{\mathscr{B}}(\mathbf{r},t) = 0, \tag{2.1b}$$

$$\nabla \times \boldsymbol{\mathscr{E}}(\mathbf{r},t) + \frac{\partial}{\partial t}\boldsymbol{\mathscr{B}}(\mathbf{r},t) = 0, \tag{2.1c}$$

$$\nabla \times \boldsymbol{\mathscr{H}}(\mathbf{r},t) - \frac{\partial}{\partial t}\boldsymbol{\mathscr{D}}(\mathbf{r},t) = \boldsymbol{\mathscr{J}}(\mathbf{r},t). \tag{2.1d}$$

Equation (2.1a) is the differential form of Gauss's law, which relates the electric displacement $\boldsymbol{\mathscr{D}}$ with the free charge density ρ. The solenoidality of the magnetic induction $\boldsymbol{\mathscr{B}}$ is enforced by Eq. (2.1b). Maxwell-Faraday's law, Eq. (2.1c), relates the electric field $\boldsymbol{\mathscr{E}}$ with $\boldsymbol{\mathscr{B}}$ and finally, Ampere's law, Eq. (2.1d), describes the connection of the magnetic field $\boldsymbol{\mathscr{H}}$, $\boldsymbol{\mathscr{D}}$ and the electric current density $\boldsymbol{\mathscr{J}}$. The real-valued electromagnetic (EM) field quantities, denoted by calligraphic bold letters, depend on the location $\mathbf{r}$ and the time t.

If we assume that there is no coupling between the electric and magnetic fields, the following constitutive relations complete the above set of equations:

$$\boldsymbol{\mathscr{D}}(\mathbf{r},t) = \epsilon_0 \boldsymbol{\mathscr{E}}(\mathbf{r},t) + \boldsymbol{\mathscr{P}}(\mathbf{r},t), \tag{2.2a}$$

$$\boldsymbol{\mathscr{H}}(\mathbf{r},t) = \mu_0^{-1} \boldsymbol{\mathscr{B}}(\mathbf{r},t) - \boldsymbol{\mathscr{M}}(\mathbf{r},t). \tag{2.2b}$$

where ϵ_0 and electric permittivity and μ_0 are the permeability of free space. The electric polarization $\boldsymbol{\mathscr{P}}$ and the magnetization $\boldsymbol{\mathscr{M}}$ account for the electric and magnetic response of the material, respectively, and are in general functions of the electric and magnetic field which can account for anisotropy, nonlinearity, and spatial and temporal nonlocality in a medium [2, 3]. Thus, the electric polarization can be expressed as

$$\boldsymbol{\mathscr{P}}(\mathbf{r},t) = \epsilon_0 \int\int \boldsymbol{\chi}_e(t,t',\mathbf{r},\mathbf{r}';\boldsymbol{\mathscr{E}})\boldsymbol{\mathscr{E}}(t,t',\mathbf{r},\mathbf{r}')d\mathbf{r}'dt', \tag{2.3}$$

where $\boldsymbol{\chi}_e$ denotes the general electric susceptibility. A similar relation can be formulated for the magnetisation with general magnetic susceptibility, $\boldsymbol{\chi}_m$. In linear, non-dispersive media, however, these relations can be substantially simplified to

$$\boldsymbol{\mathscr{P}}(\mathbf{r},t) = \epsilon_0 \boldsymbol{\chi}_e^{(1)}(\mathbf{r})\boldsymbol{\mathscr{E}}(\mathbf{r},t), \tag{2.4a}$$

$$\boldsymbol{\mathscr{M}}(\mathbf{r},t) = \mu_0^{-1} \boldsymbol{\chi}_m^{(1)}(\mathbf{r})\boldsymbol{\mathscr{B}}(\mathbf{r},t), \tag{2.4b}$$

where $\boldsymbol{\chi}_e^{(1)}$ and $\boldsymbol{\chi}_m^{(1)}$ are the linear electric and magnetic susceptibility tensors, respectively. They result in linear constitutive relations for the displacement and magnetic fields

$$\boldsymbol{\mathscr{D}}(\mathbf{r},t) = \epsilon_0 \boldsymbol{\epsilon}_r(\mathbf{r}) \boldsymbol{\mathscr{E}}(\mathbf{r},t), \tag{2.5a}$$

$$\boldsymbol{\mathscr{H}}(\mathbf{r},t) = \mu_0^{-1} \boldsymbol{\mu}_r(\mathbf{r}) \boldsymbol{\mathscr{B}}(\mathbf{r},t), \tag{2.5b}$$

with the relative permittivity tensor $\boldsymbol{\epsilon}_r(\mathbf{r}) := \boldsymbol{\chi}_e^{(1)}(\mathbf{r}) + \mathsf{I}$ and relative permeability $\boldsymbol{\mu}_r(\mathbf{r}) := \boldsymbol{\chi}_m^{(1)}(\mathbf{r}) + \mathsf{I}$. Here, I denotes the identity matrix of respective size.

Maxwell's equations can describe EM interactions with inhomogeneous structures consisting of different media. At any point on an interface S between two media with normal vector $\mathbf{n}$ pointing into the region occupied by medium 2, the electromagnetic field quantities in region 1 ($\boldsymbol{\mathscr{D}}_1$, $\boldsymbol{\mathscr{B}}_1$,$\boldsymbol{\mathscr{E}}_1$, $\boldsymbol{\mathscr{H}}_1$), are related to the quantities in region 2, denoted with subscript 2, by boundary conditions. The EM boundary conditions at this interface in the linear case read

$$(\boldsymbol{\mathscr{D}}_2 - \boldsymbol{\mathscr{D}}_1) \cdot \mathbf{n} = \sigma, \tag{2.6a}$$

$$(\boldsymbol{\mathscr{B}}_2 - \boldsymbol{\mathscr{B}}_1) \cdot \mathbf{n} = 0, \tag{2.6b}$$

$$\mathbf{n} \times (\boldsymbol{\mathscr{E}}_2 - \boldsymbol{\mathscr{E}}_1) = 0, \tag{2.6c}$$

$$\mathbf{n} \times (\boldsymbol{\mathscr{H}}_2 - \boldsymbol{\mathscr{H}}_1) = \boldsymbol{\mathscr{j}}, \tag{2.6d}$$

where σ and $\boldsymbol{\mathscr{j}}$ are the idealized free surface charge density and surface current density, respectively. The important case of vanishing σ in Eq. (2.6a) implies continuity of the normal component of the displacement field across the interfaces of different media. The tangential component of $\boldsymbol{\mathscr{E}}$ is always continuous. If the to media additionally have different permittivity, $\epsilon_1 \neq \epsilon_2$, the normal component of $\boldsymbol{\mathscr{E}}$ and tangential components of $\boldsymbol{\mathscr{D}}$ are discontinuous across S.

A direct consequence of ME is the *continuity relation* for free charge density and current density ρ and $\boldsymbol{\mathscr{J}}$, respectively. The divergence of Eq. (2.1d) together with Eq. (2.1a) yields

$$\frac{\partial}{\partial t}\rho(\mathbf{r},t) + \nabla \cdot \boldsymbol{\mathscr{J}}(\mathbf{r},t) = 0. \tag{2.7}$$

An important aspect of ME (2.1) is revealed if harmonic time-dependence $e^{-i\omega t}$ of all fields is assumed. Specifically, for a fixed angular frequency ω, each field quantity $\mathscr{f}(\mathbf{r},t)$ is written as

$$\mathscr{f}(\mathbf{r},t) = f(\mathbf{r},\omega)e^{-i\omega t} + \text{c.c.} \tag{2.8}$$

with a complex, time independent amplitude $f(\mathbf{r},\omega)$. It should be noted that there are different conventions regarding Eq. (2.8) as it can involve an additional factor of $\frac{1}{2}$. The harmonic time-dependence allows the replacement of the partial temporal derivatives in Eq. (2.1) with $-i\omega$ and the formulation of ME in the frequency-domain

$$\nabla \cdot \mathbf{D}(\mathbf{r},\omega) = \rho(\mathbf{r},\omega) \tag{2.9a}$$

$$\nabla \cdot \mathbf{B}(\mathbf{r},\omega) = 0 \tag{2.9b}$$

$$\nabla \times \mathbf{E}(\mathbf{r},\omega) - i\omega\mathbf{B}(\mathbf{r},\omega) = 0 \tag{2.9c}$$

$$\nabla \times \mathbf{H}(\mathbf{r},\omega) + i\omega\mathbf{D}(\mathbf{r},\omega) = \mathbf{J}(\mathbf{r},\omega) \tag{2.9d}$$

It should be noted that, without electromagnetic sources, Eqs. (2.9a) and (2.9b) are not independent but can be obtained by taking the divergence of Eqs. (2.9c) and (2.9d), respectively. Given a field quantity for all frequencies ω one can recover its behaviour in the time-domain by the inverse Fourier transform. The electric field is given by the linear superposition of the single frequency solutions $\mathbf{E}(\mathbf{r},\omega)$ of Eq. (2.9)

$$\boldsymbol{\mathscr{E}}(\mathbf{r},t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbf{E}(\mathbf{r},\omega)e^{-i\omega t}d\omega. \tag{2.10}$$

A special kind of solution of ME in a homogeneous, isotropic ($\boldsymbol{\epsilon}_r(\mathbf{r}) = \epsilon_r\mathsf{I}, \boldsymbol{\mu}_r(\mathbf{r}) = \mu_r\mathsf{I}$) medium without sources, $\rho(\mathbf{r},t) = 0$ and $\boldsymbol{\mathscr{J}}(\mathbf{r},t) = 0$, is the transverse plane wave

$$\boldsymbol{\mathscr{E}}(\mathbf{r},t) = \mathbf{E}(\omega)e^{i(k\mathbf{n}\cdot\mathbf{r}-\omega t)} + \text{c.c.} \tag{2.11}$$

where $\mathbf{n}$ is the unit vector along the propagation direction and k is called the wave number. To show this, Eqs. (2.9c) and (2.9d) can be combined to obtain the Helmholtz wave equation in frequency-domain without sources

$$\Delta\mathbf{E}(\mathbf{r},\omega) + \omega^2\mu\epsilon\mathbf{E}(\mathbf{r},\omega) = 0, \tag{2.12a}$$

$$\Delta\mathbf{B}(\mathbf{r},\omega) + \omega^2\mu\epsilon\mathbf{B}(\mathbf{r},\omega) = 0 \tag{2.12b}$$

where Δ denotes the Laplace operator: $\Delta\mathbf{F} := \nabla(\nabla\cdot\mathbf{F}) - \nabla\times(\nabla\times\mathbf{F})$ for any vector field $\mathbf{F}$ of $\mathbf{r}$. Inserting the definition of the plane wave Eq. (2.11) into the Helmholtz equation (2.12) yields the expression for the wavenumber

$$k = \sqrt{\mu_0\epsilon_0\mu_r\epsilon_r}\omega = \sqrt{\mu_0\epsilon_0}n\omega = \omega\frac{n}{c} = k_0 n. \tag{2.13}$$

Hereby, $n = \sqrt{\mu_r\epsilon_r}$ denotes the refractive index of the medium and $c = \frac{1}{\sqrt{\mu_0\epsilon_0}}$ is the speed of light in vacuum. Moreover, $k_0 = \frac{\omega}{c}$ is the free space wavenumber.

The material electrical properties, namely the electric and magnetic susceptibility, of all real media exhibit dispersion, that is they vary with frequency ω [2]. This is illustrated by considering the motion of an electron with negative unit charge $-e$ and mass m around a fixed positive charge (e.g. atomic core) in an isotropic medium. Let us assume that the electron at position $\mathbf{r}$ is bound by a harmonic force with binding frequency ω_b and additionally excited by a harmonic electric field $\boldsymbol{\mathscr{E}}(\mathbf{r},t) = \mathbf{E}(\mathbf{r})e^{-i\omega t}$. This can be described by the driven harmonic oscillator model

$$m\left(\ddot{\mathbf{r}} + \gamma\dot{\mathbf{r}} + \omega_b^2\mathbf{r}\right) = -e\boldsymbol{\mathscr{E}}(\mathbf{r},t) \tag{2.14}$$

where γ is a phenomenological damping factor. Assuming only small deviations from the average electron position one can find the electric dipole moment contribution by each electron to be

$$\mathbf{p} = -e\mathbf{r} = \frac{e^2}{m\left(\omega_b^2 - \omega - i\omega\gamma\right)}\mathbf{E} \tag{2.15}$$

Assuming further that the medium consists of N molecules per unit volume and f_j electrons per molecule with binding frequency ω_j and damping factor γ_j, $j = 1, \ldots, N_j$, one finds for the linear electric susceptibility

$$\chi_e(\omega) = \frac{Ne^2}{\epsilon_0 m}\sum_{j=1}^{N_j}\frac{f_j}{\omega_j^2 - \omega^2 - i\omega\gamma_j}. \tag{2.16}$$

Denoting $\frac{Ne^2}{\epsilon_0 m} := \omega_p^2$ as the plasma frequency and also including the contribution of free electrons, i.e. electrons with vanishing binding frequency $\omega_b = 0$, one obtains the Drude-Lorentz model for the relative permittivity $\epsilon_r(\omega) = 1 + \chi_e(\omega)$

$$\epsilon_r(\omega) = 1 + i\frac{\omega_p^2 f_0}{\omega(\gamma_0 - i\omega)} + \sum_{j=1}^{N_j}\frac{\omega_p^2 f_j}{\omega_j^2 - \omega^2 - i\omega\gamma_j} \tag{2.17}$$

where f_0 is the number of free electrons per molecule and γ_0 their damping factor. This is a general and widely used model for the electrical permittivity of metals and semiconductors. In this derivation, magnetic effects have been ignored and in fact, natural materials do not exhibit magnetism at optical frequencies. Therefore, only nonmagnetic materials ($\mu_r = 1$) are considered throughout this thesis.

For the investigation of ME it is often useful to formulate them in curvilinear coordinate systems [4, 5]. Depending on the structure under consideration, ME can be simplified significantly if a suitable coordinate system is chosen. For example the scattering of light from a spherical structure can most elegantly be described in a spherical coordinate sys-

tem [6].

Let $Or^1r^2r^3 = Oxyz$ denote the Cartesian coordinate system and let $O\bar{r}^1\bar{r}^2\bar{r}^3$ be an arbitrary curvilinear coordinate system which is related to $Or^1r^2r^3$ by the coordinate transform action

$$r^\alpha = r^\alpha(\bar{r}^1, \bar{r}^2, \bar{r}^3), \quad \alpha = 1, 2, 3. \tag{2.18}$$

In a structure containing a distribution of nonmagnetic, isotropic media free of sources Faraday's Law (2.9c) and Ampere's Law (2.9d) in covariant form then read

$$\sum_{l,m=1}^{3} \zeta^{klm} \frac{\partial}{\partial \bar{r}^l} \bar{E}_m(\omega, \bar{\mathbf{r}}) = ik_0 \sum_{l=1}^{3} |\bar{\mathbf{g}}(\bar{\mathbf{r}})|^{-\frac{1}{2}} \bar{g}^{kl}(\bar{\mathbf{r}}) \bar{H}_l(\omega, \bar{\mathbf{r}}), \tag{2.19a}$$

$$\sum_{l,m=1}^{3} \zeta^{klm} \frac{\partial}{\partial \bar{r}^l} \bar{H}_m(\omega, \bar{\mathbf{r}}) = -ik_0 \sum_{l=1}^{3} |\bar{\mathbf{g}}(\bar{\mathbf{r}})|^{-\frac{1}{2}} \bar{g}^{kl}(\bar{\mathbf{r}}) \epsilon(\mathbf{r}) \bar{E}_l(\omega, \bar{\mathbf{r}}) \tag{2.19b}$$

where $\bar{E}_k$ and $\bar{H}_k$ are field components in the curvilinear coordinate system. Here, ζ denotes the totally antisymmetric Levi-Civita tensor and $\bar{\mathbf{g}}$ is the metric tensor associated with $O\bar{r}^1\bar{r}^2\bar{r}^3$ with the components

$$\bar{g}^{kl}(\bar{\mathbf{r}}) = \sum_{n=1}^{3} \frac{\partial \bar{r}^k}{\partial r^n} \frac{\partial \bar{r}^l}{\partial r^n} \tag{2.20}$$

and determinant $|\bar{\mathbf{g}}(\bar{\mathbf{r}})|$. One can identify effective relative permittivity and magnetic susceptibility tensors

$$\bar{\epsilon}^{kl}(\bar{\mathbf{r}}) = \epsilon(\mathbf{r}) \, |\bar{\mathbf{g}}(\bar{\mathbf{r}})|^{-\frac{1}{2}} \bar{g}^{kl}(\bar{\mathbf{r}}), \tag{2.21a}$$

$$\bar{\mu}^{kl}(\bar{\mathbf{r}}) = |\bar{\mathbf{g}}(\bar{\mathbf{r}})|^{-\frac{1}{2}} \bar{g}^{kl}(\bar{\mathbf{r}}) \tag{2.21b}$$

that have been induced by the transformation into a curvilinear coordinate system. Changing the coordinate system is hence equivalent to a change of the material properties. This principle is at the heart of the theory of transformation optics [7].

2.3 Nomenclature and theoretical aspects of periodic structures: diffraction gratings

Periodicity is an important property of many EM structures as it allows the coherent scattering of light and consequently enhanced device functionality. A random arrangement of N scatterers results in incoherent scattering of light, the total intensity of scattered light being proportional to N and usually independent of the considered direction [2]. By contrast, in the case of a periodic arrangement of N scatterers the intensity of scattered light

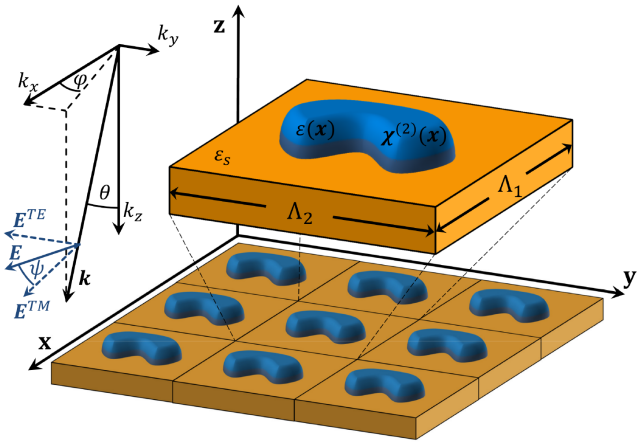

Figure 2.1: Generic two-dimensional grating (bottom) with periods Λ_1 and Λ_2 on a homogeneous substrate with permittivity ϵ_s and close-up of its unit cell. The cover material is vacuum.

is proportional to N^2, if the scattering direction is along specific directions. This is due to the coherent superposition of individual scattering amplitudes [2, chapter 10]. Therefore, periodicity is one of the defining feature of different optical devices.

Before some of the most important applications of diffraction gratings and other periodic optical structures are presented in Section 2.6, a few definitions and theoretical concepts about periodic structures in the context of electromagnetics will be given and discussed. To this end, this section borrows concepts from the theory of periodic media [8, 9] and it is assumed that all involved media are nonmagnetic ($\mu_r(\mathbf{r}) = 1$), linear and isotropic ($\boldsymbol{\epsilon}_r(\mathbf{r}) = \epsilon(\mathbf{r})\mathsf{I}$).

A generic two-dimensionally periodic grating structure is depicted in Fig. 2.1. The structure possesses a discrete translational invariance in the x-y-plane but is arbitrary in the z-direction:

$$\epsilon_r(\mathbf{r}) = \epsilon_r(\mathbf{r} + \mathbf{R}) \tag{2.22}$$

for all lattice vectors $\mathbf{R} \in \mathscr{R} := \mathscr{R}(\boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2) := \{n_1\boldsymbol{\Lambda}_1 + n_2\boldsymbol{\Lambda}_2 \,:\, n_i \in \mathbb{Z}\}$. The two linearly independent vectors $\boldsymbol{\Lambda}_i$ are called primitive lattice vectors as their discrete linear combinations form the set of lattice vectors $\mathscr{R}$. The *unit cell* is the smallest area (volume in 3D) of the direct space such that it generates the direct lattice by translation with the lattice vectors.

Gratings are conceptually similar to 2D photonic crystals, as per Fig. 2.2(b), in that both are periodic in two dimensions, with one distinct difference: the characteristic length of 2D photonic crystals is comparable to the optical wavelengths, whereas this constraint does not hold on the case of diffraction gratings. Every optical structure fulfilling this criterion will be called *diffraction grating*, or simply *grating* throughout this work. This broader

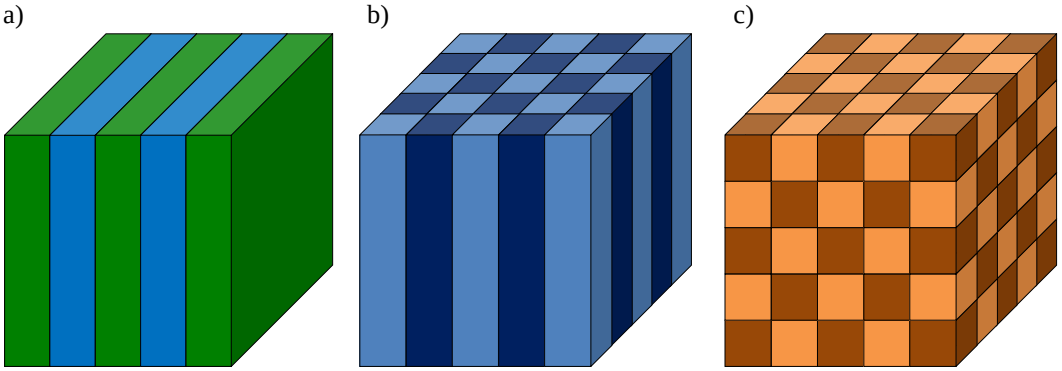

Figure 2.2: Schematic representation of one-, two- and three-dimensional photonic crystals. Colours represent media with different permittivities (adapted from Ref. [8]).

definition of the term grating is fruitful as it covers very general structures, e.g. periodic metamaterials and generalizes their theoretical and numerical treatment, and at the same time covers the traditional, more specific terms *binary gratings, crossed* or *holographic gratings*.

An important concept derived from this periodicity is that of the *reciprocal lattice*. By applying a Fourier transform to $\epsilon(\mathbf{r})$ and using the periodicity of $\epsilon(\mathbf{r})$ one can find that for each z, $\epsilon(\mathbf{r})$ is given as two-dimensional Fourier series

$$\epsilon_r(x, y, z) = \sum_{\mathbf{G}\in\mathscr{G}} \tilde{\epsilon}_r(\mathbf{G}, z) e^{i\mathbf{G}\cdot\mathbf{r}_\parallel} \tag{2.23}$$

where $\mathscr{G} = \mathscr{G}(\mathbf{\Lambda}_1, \mathbf{\Lambda}_2) := \{\mathbf{G}$ such that for any $\mathbf{R} \in \mathscr{R}$ there is a n with $\mathbf{G}\cdot\mathbf{R} = 2n\pi\}$ denotes the set of reciprocal lattice vectors. Interestingly, one can show that the reciprocal lattice of $\mathscr{G}$ is again the direct lattice $\mathscr{R}$, which emphasises their dual nature.

The reciprocal lattice allows the definition of the *first Brillouin zone* $\mathscr{B}$, a non redundant domain in the reciprocal space. It is defined as the domain in the reciprocal space for which any point is closer to the origin than to any other reciprocal lattice vector:

$$\mathscr{B} := \{\mathbf{k} : \|\mathbf{k}\| \leq \|\mathbf{k} + \mathbf{G}\| \text{ for all } \mathbf{G} \in \mathscr{G}\} \tag{2.24}$$

The elements of $\mathscr{B}$ are non redundant in the sense, that there are no two distinct vectors $\mathbf{P} \neq \mathbf{Q}$ in the interior of $\mathscr{B}$ such that $\mathbf{P} - \mathbf{Q} = \mathbf{G}$ for any $\mathbf{G} \in \mathscr{G}$. The direct and reciprocal lattice and the first Brillouin zone for a triangular lattice are illustrated in Fig. 2.3.

From this starting point one can already characterize an important property of the EM fields of a grating structure. It can be shown (see Appendix A) that all eigensolutions of ME for a grating are pseudo-periodic. That is for each in-plane wavevector $\mathbf{k}_\parallel$ one can write a

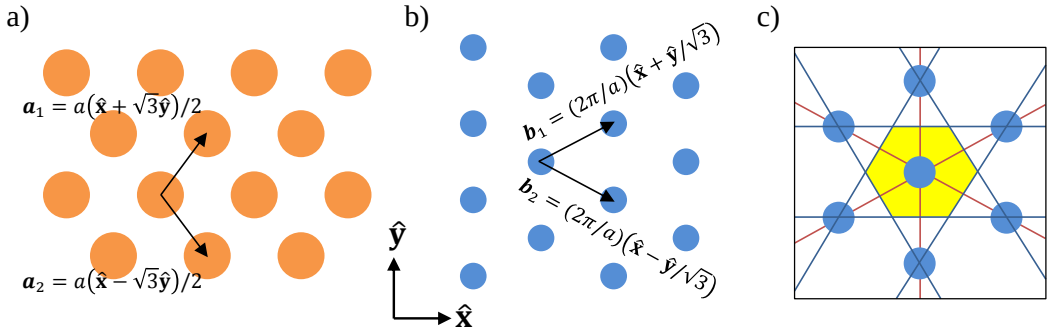

Figure 2.3: A triangular lattice in real and reciprocal space (adapted from Ref. [8]). a) Network of lattice points in real space. b) Representation of the reciprocal lattice in k-space. c) The first Brillouin zone is a hexagon centred at the origin.

corresponding solution $\mathbf{E}_{\mathbf{k}_\parallel}$ of Eq. (2.9)

$$\mathbf{E}_{\mathbf{k}_\parallel}(\mathbf{r}, \omega) = \mathbf{u}_{\mathbf{k}_\parallel}(\mathbf{r}_\parallel, z, \omega) e^{i\mathbf{k}_\parallel \cdot \mathbf{r}} \tag{2.25}$$

as the product of a periodic function $\mathbf{u}_{\mathbf{k}_\parallel}(\mathbf{r}_\parallel, z, \omega)$ and a plane wave phase shift term $e^{i\mathbf{k}_\parallel \cdot \mathbf{r}}$. It should be noted that $\mathbf{u}_{\mathbf{k}_\parallel}(\mathbf{r}_\parallel, z, \omega)$ is periodic only in the x-y-plane. Its longitudinal z-dependence is still arbitrary. This fundamental result is commonly referred to as Bloch's Theorem. It has been established first in the context of quantum mechanics [10] and solid-state physics [11], where it states that the wave function of a particle in a periodic potential is pseudo-periodic. It is the starting point for the theoretical investigation of gratings and many numerical solution methods for gratings, see Section 2.5.

For a given plane wave which is incident from a homogeneous cover region with refractive index n_c onto a grating structure, as per Fig. 2.1, with wavevector $\tilde{\mathbf{k}} = (k_x, k_y, -k_z)$ and amplitude $\tilde{\mathbf{E}}_0'$, Bloch's Theorem allows to decompose the EM fields in the cover into a pseudo-Fourier series

$$\mathbf{E}^{(1)}(\omega, \mathbf{x}) = \tilde{\mathbf{E}}_0 e^{i\tilde{\mathbf{k}} \cdot \mathbf{x}} + \sum_{\mathbf{G} \in \mathscr{G}} \mathbf{E}_{\mathbf{G}}'^{(1)} e^{i\mathbf{k}_\mathbf{G} \cdot \mathbf{x}}. \tag{2.26}$$

The single plane wave terms in the infinite sum are called *diffraction orders* and they constitute the reflected field. Their respective wavevectors are given by

$$\mathbf{k}_{\mathbf{G},\parallel} = \tilde{\mathbf{k}}_\parallel + \mathbf{G}_\parallel \qquad \text{and} \qquad \mathbf{k}_{\mathbf{G},z} = \sqrt{\omega^2 n_c^2 c^{-2} - \left\|\mathbf{k}_{\mathbf{G},\parallel}\right\|^2}. \tag{2.27}$$

The latter formula represents the dispersion relation of a plane wave propagating in the cover medium. For small $\left\|\mathbf{G}_\parallel\right\|$, $\mathbf{k}_{\mathbf{G},z}$ is a real number, which yields a small number of *propagating* diffraction orders. Their respective intensity relative to the incident intensity is called *diffraction efficiency*. For $\left\|\mathbf{G}_\parallel\right\|$ large enough, $\mathbf{k}_{\mathbf{G},z}$ becomes imaginary and yields

an *evanescent* plane wave. Such evanescent plane waves do not carry energy, but they contribute to the near-field of the grating structure.

An interesting phenomenon for changing ω appears when a diffraction order undergoes the change from propagating to evanescent, i.e. $\mathbf{k}_{\mathbf{G},z} = 0$. This transition can be due to changing of the incident angle or wavelength and leads to a redistribution of power between the remaining propagating orders. This phenomenon which has been studied in Ref. [12] and is known as cut-off or Wood's anomaly. A decomposition analogous to Eq. (2.26) can be done in the homogeneous substrate region.

2.4 Introduction to nonlinear optics

Nonlinear optics is the study of the interaction of intense light with media in which the electric polarization density $\boldsymbol{\mathscr{P}}$ depends on the electric field $\boldsymbol{\mathscr{E}}$ in a nonlinear fashion. Sufficiently intense light, e.g. generated by a strong laser source, can modify the optical properties of a material system itself resulting in such a nonlinear behaviour. This is an underlying principle for many applications and an introduction into its mathematical description is given in what follows.

The nonlinear polarization can be described by a Taylor series

$$\begin{aligned}\boldsymbol{\mathscr{P}}(\mathbf{r},t) =& \epsilon_0 \boldsymbol{\chi}^{(1)} \boldsymbol{\mathscr{E}}(\mathbf{r},t) + \epsilon_0 \boldsymbol{\chi}^{(2)} : \boldsymbol{\mathscr{E}}(\mathbf{r},t)\boldsymbol{\mathscr{E}}(\mathbf{r},t) + \epsilon_0 \boldsymbol{\chi}^{(3)} \vdots \boldsymbol{\mathscr{E}}(\mathbf{r},t)\boldsymbol{\mathscr{E}}(\mathbf{r},t)\boldsymbol{\mathscr{E}}(\mathbf{r},t) + \dots \\ =& \boldsymbol{\mathscr{P}}^{(1)}(\mathbf{r},t) + \boldsymbol{\mathscr{P}}^{(2)}(\mathbf{r},t) + \boldsymbol{\mathscr{P}}^{(3)}(\mathbf{r},t) + \dots \end{aligned} \tag{2.28}$$

where $\boldsymbol{\chi}^{(1)}$ is the linear electric susceptibility, and $\boldsymbol{\chi}^{(2)}$ and $\boldsymbol{\chi}^{(3)}$ are the second- and third-order susceptibilities, respectively, and no time-dependence of the susceptibility values was assumed. The second- and third-order polarization components $\boldsymbol{\mathscr{P}}^{(2)}$ and $\boldsymbol{\mathscr{P}}^{(3)}$, respectively, are typically orders of magnitude smaller than the linear polarization $\boldsymbol{\mathscr{P}}^{(1)}$ due to the minute values of $\boldsymbol{\chi}^{(2)}$ and $\boldsymbol{\chi}^{(3)}$ in natural materials compared to $\chi^{(1)}$. Hence the linear approximation $\boldsymbol{\mathscr{P}}(\mathbf{r},t) = \boldsymbol{\mathscr{P}}^{(1)}(\mathbf{r},t)$, which underlies the theory of linear optics, is sufficient when the intensity of $\boldsymbol{\mathscr{E}}(\mathbf{r},t)$ is small. For increasing intensity, however, the higher-order contributions become relevant. Amongst the first experimental observations of nonlinear optical phenomena was that of second-harmonic generation in a quartz crystal under strong laser field by Franken et al. [13]. In fact, given high enough intensity of the electric field, all dielectrics, even vacuum, become nonlinear [14].

Second-harmonic generation (SHG), a special case of sum-frequency generation (SFG), is one of the most important nonlinear optical processes. It occurs when two photons with energy $E = \hbar\omega$ combine to form a photon with twice their energy $2E = 2\hbar\omega$. A schematic of the involved energy levels and a generic configuration of SHG is shown in Fig. 2.4(a) and (b), respectively. In a first approximation, SHG only occurs in the bulk of non-centrosymmetric media, that is in media whose crystal structure is not invariant to the

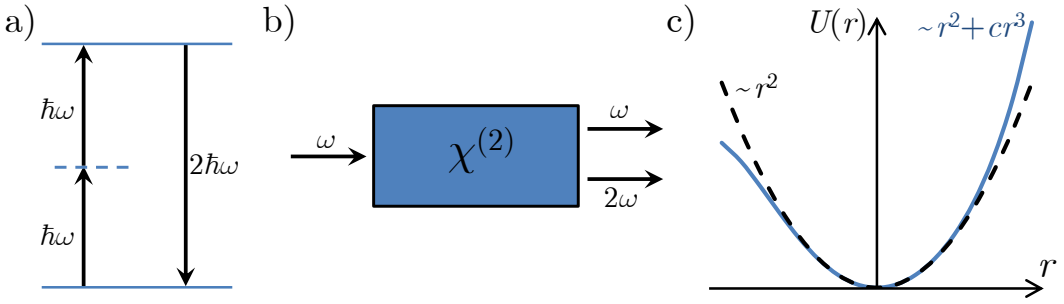

Figure 2.4: Theoretical aspects of second harmonic generation (adapted from Ref. [15]). a) Energy level diagram of photons for SHG. b) Schematic of frequency doubling by a nonlinear optical material. c) Potential energy Function for a centrosymmetric (dashed) and a non-centrosymmetric medium (solid blue).

inversion symmetry transformation.

A fundamental model for the second-order nonlinear susceptibility in such a medium can be derived by considering the potential energy function $U(r)$ of a non-centrosymmetric material. Due to the lack of inversion symmetry of the material, U is asymmetric and the simplest example is a cubic dependence of U

$$U(r) = \int F(r)dr = -m\left(\frac{1}{2}\omega^2 r^2 + \frac{1}{3}ar^3\right), \tag{2.29}$$

where the parameter a characterizes the strength of the anharmonicity (nonlinearity) of the system. This is illustrated in Fig. 2.4(c). Not incidentally, Eq. (2.29) can be seen as the second and third terms of a Taylor expansion of $U(r)$ in the displacement r of the electron from its equilibrium position. This potential gives rise to an anharmonic restoring force

$$F(r) = -m(\omega^2 r - ar^2) \tag{2.30}$$

in the oscillator model Eq. (2.14) and yields

$$m\left(\ddot{r} + \gamma\dot{r} + \omega^2 r + ar^2\right) = -eE(r)e^{-i\omega t} + \text{c.c.} \tag{2.31}$$

By solving this ordinary differential equation (ODE) one obtains relations for the nonlinear polarization

$$P(2\omega) = \epsilon_0 \chi^{(2)}(\omega)E^2(\omega) \tag{2.32}$$

and the expression for the second order susceptibility

$$\chi^{(2)}(\omega) = \frac{\epsilon_0^2 ma}{N^2 e^3}\chi^{(1)}(2\omega)\left(\chi^{(1)}(\omega)\right)^2. \tag{2.33}$$

Similar expressions for sum- and difference-frequency generation can be obtained the same way, see Ref. [15, chapter 1]. Note that a scalar approximation of the electric field was made in Eq. (2.31) and the specific tensorial structure of $\boldsymbol{\chi}^{(2)}$ has not been considered here. The most general form of the second-harmonic polarization in non-centrosymmetric media incorporating the vectorial nature of $\mathbf{P}$ and $\mathbf{E}$ is given by

$$P_i(2\omega) = \epsilon_0 \sum_{j,k} \chi^{(2)}_{ijk}(\omega) E_j(\omega) E_k(\omega), \tag{2.34}$$

where the specific structure of the complex $3 \times 3 \times 3$ tensor $\boldsymbol{\chi}^{(2)}$ is determined by the symmetry properties of the material [16, chapter 2]. Important non-centrosymmetric materials with high second-order susceptibility are GaAs, ZnO, $LiNO_3$ and $LiIO_3$. In a centrosymmetric material, on the other hand, $\boldsymbol{\chi}^{(2)}$ vanishes in the above presented approximation, as the potential U is symmetric in media with inversion symmetry, i.e. $a = 0$.

By including mbr^3, the next term in the Taylor expansion, into the restoring force, one can derive third-order nonlinear optical effects from Eq. (2.31). The first example of a third-order nonlinear process is third-harmonic generation (THG), the generation of EM fields with frequency 3ω from an incident field with frequency ω, which is characterized by a nonlinear polarization

$$P_i^{\mathrm{NL}}(3\omega) = \epsilon_0 \sum_{jkl} \chi^{(3)}_{ijkl}(\omega + \omega + \omega) E_j(\omega) E_k(\omega) E_l(\omega). \tag{2.35}$$

The second example is the optical Kerr effect and amounts to a field intensity dependent change of the index of refraction

$$n = n(\mathbf{E}) = n_0 + 2\bar{n}_2 |E(\omega)|^2 \tag{2.36}$$

where n_0 denotes the index of refraction in the linear regime and $\bar{n}_2$ is called the nonlinear-index coefficient. This can also be expressed in terms of a nonlinear polarization

$$P^{\mathrm{NL}}(\omega) = 3\epsilon_0 \chi^{(3)}(\omega + \omega - \omega) |E(\omega)|^2 E(\omega) \tag{2.37}$$

where for simplicity, tensor indices of $\chi^{(3)}$ are suppressed and the field $\mathbf{E} = \mathbf{e}_x E$ is assumed to be linearly polarized. The nonlinear-index coefficient is hence related to $\chi^{(3)}$ via $\bar{n}_2 = \frac{3\chi^{(3)}}{4n_0}$.

Despite the vanishing second-order nonlinearity $\boldsymbol{\chi}^{(2)}$ in the bulk of centrosymmetric materials, SHG can occur in such materials in two ways [17, 18]: *i*) The inversion symmetry of centrosymmetric materials is broken at their interface, yielding a locally confined surface

SH polarization

$$\mathbf{P}^{\mathrm{NL}}(2\omega) = \epsilon_0 \boldsymbol{\chi}_s^{(2)}(\omega) : \mathbf{E}(\omega)\mathbf{E}(\omega) \tag{2.38}$$

which is only nonvanishing at the surface of the material. The intensity of the generated surface SH is typically much smaller than in bulk of non-centrosymmetric materials. *ii*) By taking into account the nonlocal contributions to the nonlinear polarization, that is the dependence of $\mathbf{P}^{\mathrm{NL}}$ on spatial derivatives of $\mathbf{E}$, a nonlocal nonlinear source polarization for SHG in the bulk of centrosymmetric materials can be formulated:

$$\mathbf{P}^{\mathrm{NL}}(2\omega) = \epsilon_0 \boldsymbol{\chi}_q^{(2)}(\omega) \vdots \mathbf{E}(\omega)\nabla\mathbf{E}(\omega) \tag{2.39}$$

where the explicit denotation of the 81 tensor components of $\boldsymbol{\chi}_q^{(2)}$ has been dropped. It should be noted that noble metals, most liquids and gases, and - very important for the emerging field of silicon photonics - silicon are centrosymmetric.

From the above discussion it can be seen that intense electric fields are necessary in order to achieve strong nonlinear interactions, due to the generally small values of the nonlinear susceptibilities $\chi^{(2)}$ and $\chi^{(3)}$. An important concept to achieve a strong nonlinear optical response is phase-matching: In the case of SFG and collinear plane wave propagation, the intensity of the generated field at frequency $\omega_3 = \omega_1 + \omega_2$ is given by

$$I = I_{\max} \frac{\sin^2(\Delta k L/2)}{(\Delta k L/2)^2} \tag{2.40}$$

where $I_{\max}$ denotes the total incident field intensity and L is the propagation length through the nonlinear medium. The important quantity $\Delta k = k_1 + k_2 - k_3$ is called the wavevector mismatch of the beams with wavenumbers k_i. This relation implies *i)* a drastic decrease of the achieved intensity at the generated frequency ω_3 if the wavevectors are not matched

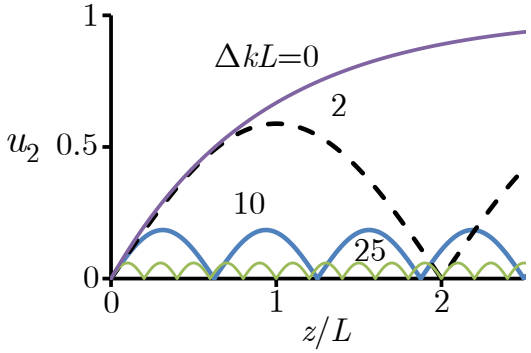

Figure 2.5: Effect of wavevector mismatch on the efficiency of second-harmonic generation (adapted from Ref. [15]).

($\Delta k \neq 0$) and *ii)* a periodic dependence of the generated intensity on the effective propagation length L if $\Delta k \neq 0$. Both effects are illustrated in Fig. 2.5 for the special case of SHG where u_2 denotes the real, normalized field amplitude at the SH, ΔkL is the normalized phase mismatch parameter. Note that most lossless materials exhibit normal dispersion, i.e. their refractive index monotonically increases for increasing frequency, which prohibits direct phase matching. There are three ways to still achieve exact phase matching: *i)* Near an absorption feature of a dispersion relation of the material, it exhibits anomalous dispersion. *ii)* Birefringence, which is the dependence of refractive index on the polarization and direction of propagation, can be used to fulfil $\Delta k = 0$ in certain anisotropic materials. *iii)* A third approach is quasi- phase-matching by periodically poling the medium.

2.5 Overview of computational methods in electromagnetics

The study of the interaction of electromagnetic waves with matter has spawned a large variety of methods to analytically or numerically solve Maxwell's equations. The results obtained from those methods are invaluable for understanding, validating and predicting experimental results and for the design process of optical devices. Before numerical methods for solving Maxwell's equation are discussed, it should be noted that there has been and still is some interest in analytical and semi-analytical solution strategies for electromagnetic problems. These solution-approaches can be highly efficient and elegant. Their drawback is that they practically only work for a narrow range of problems and under limiting assumptions. However, the theoretical insight gained by analytical investigation of ME is valuable and such analytical methods often form the basis of numerical methods or can be used to obtain reference solutions for validation of the computer implementation of a numerical method.

Moreover, the transition between an analytical method and a numerical method can be seamless, as in the example of geometrical optics [19, 20]. The analytical investigation of the short-wavelength limit of the wave-equation yields geometrical optics, which allows one to describe large systems of paraxial optical elements (lenses, prisms, mirrors) in a simple and very descriptive matrix formalism. However, the regime, where an approximation is valid is limited: geometrical optics fails when diffraction or interference play a role, i.e. when the wavelength is not much smaller than characteristic size of a device.

The focus of this work lies on rigorous numerical methods for electromagnetics, in particular rigorous numerical methods to describe linear and nonlinear interaction of light with diffraction gratings comprising conventional bulk and novel 2D materials. Before introducing the formalism of numerical methods for periodic structures, a short classification of rigorous numerical methods will be given now. Such an overview is important because, naturally, no method can be universally used for any structure and setting; Each rigorous method has a range of applications for which it is particularly well or ill suited. Knowledge

of their properties is necessary to select an appropriate method to model a certain structure at hand with high accuracy at minimal computational cost.

2.5.1 General-purpose numerical methods

The most fundamental distinction between numerical methods for ME is between time- and frequency-domain methods. *i*) A time-domain methods solves the time-dependent ME (2.1) in a time-stepping manner, i.e. a given initial configuration of the EM fields is advanced from one time step to another. The numerical quantities are the real-valued time dependent EM fields $\mathscr{E}$ and $\mathscr{H}$. A direct advantage of time-domain methods is that they describe the transient behaviour of the EM fields. *ii*) Frequency domain methods solve ME in the frequency-domain, Eqs. (2.9), for a single given frequency, ω. The numerical quantities are the complex field amplitudes $\mathbf{E}$ and $\mathbf{H}$ consisting of their real amplitude and phase. Frequency domain methods hence immediately calculate the time-harmonic behaviour of a system, which may be hard to achieve by simulating the transient behaviour with time-domain methods.

Note that there are also methods that do not solve ME in the differential form (Eq. (2.1) or Eq. (2.9)) but solve their integral form. Such methods include the method of moments (also known as boundary element method [21]) and the discrete dipole approximation [22]. Their use can be especially beneficial for scattering and radiation problems in unbounded domains.

Having distinguished between time- and frequency-domain methods, another important distinction is the spatial discretisation employed by a method: *i*) Finite difference (FD) discretisation employs a rectangular grid of points to discretise the spatial computational domain. In the case of computational electromagnetics the Yee lattice [23], a staggered grid in which the unknowns $\mathscr{E}_\alpha$ are located between the values of $\mathscr{H}_\beta$, is used to simplify the calculation of the curl-operator and grant numerical robustness. Finite difference discretisation is conceptually simple and can be highly efficient and accurate for grid-aligned structures, but struggles with oblique and rounded shapes. *ii*) Finite element (FE) methods are based on calculus of variations and their discretisation typically employs a structured tetrahedral (triangular in 2D) mesh. Structured meshes allow the discretisation of geometrically arbitrary, non-aligned and rounded structures with high accuracy at the cost of slightly higher conceptual complexity. A less commonly used spatial discretisation scheme is provided by the finite volume [24] discretisation. One problem from which all space domain methods suffer is the finite size of the discretised computational domain for radiating or scattering applications, which can result in artificial reflection from the boundaries of the discretised domain. Perfectly matched layers (PML), which are artificially introduced absorbing material layers at the boundary, can be employed to remedy such unwanted effects.

A popular method in computational electromagnetics is the finite-difference time-

domain method (FDTD, [25]). Its most compelling features are its efficiency, simplicity and capability to solve electromagnetic problems coming from a wide range of applications. The efficiency of FDTD can be attributed to the use of an explicit time discretisation scheme and the structured Yee lattice. Although the time step size of explicit methods is limited by the Courant-Friedrichs-Lewy (CFL) stability condition, each FDTD time step can be computed very efficiently as it only requires an explicit update of grid variables from neighbouring values which does not involve the solution of a potentially large linear system of equations. There are numerous commercially or freely available FDTD implementations [26–28], as well as countless hand-tailored implementations developed and used in academic institutions and industrial research and development institutions.

For certain classes of structures, there are more specialized algorithms such as the eigenmode expansion method [29] or the split-step Fourier method [30] for light propagation in optical fibres and waveguides. Other specialised numerical methods are the multiple scattering method [31], a series expansion method in the frequency-domain that is tailored to calculate interaction of light with clusters of particles [6] or rods [32]. Applying these methods to the specific domain or geometry, for which they are developed, allows one to simulate more accurately and efficiently the EM interaction with these structures. This makes highly specialised methods superior to general purpose methods when applied to these particular geometries at the cost of the additional theoretical knowledge that has to be invested in their derivation.

Another specific domain of applications, which highly benefits from special numerical analysis, are periodic structures, in particular diffraction gratings, which are the focus of this thesis. An overview of specialized numerical methods for diffraction gratings is presented in the next section.

2.5.2 Numerical methods for diffraction gratings

Although there are ways to simulate diffraction gratings with general time-domain methods like FDTD or finite-element time-domain (FETD), on many occasions it is more beneficial to use more specialized algorithms, particularly frequency-domain methods. The major problem of time-domain methods for gratings is, amongst others, that the frequency dependent Bloch phase shift, i.e. the phase shift $e^{i\mathbf{k}_\parallel \cdot \Lambda}$ from one unit cell to the next, given by the tangential component of the incident wavevector, can not be directly modelled by time-domain methods without major modifications, such as the split field formulation of FDTD (see Chapter 13 of Ref. [25] and Chapter 9 of Ref. [33]). The direct approach of modelling periodic structures in a time-domain method is to not explicitly incorporate the phase shift, but instead simulate a structure consisting of several identical unit cells. Although this approach is conceptually simple, it is mostly ruled out due to its high computational demand.

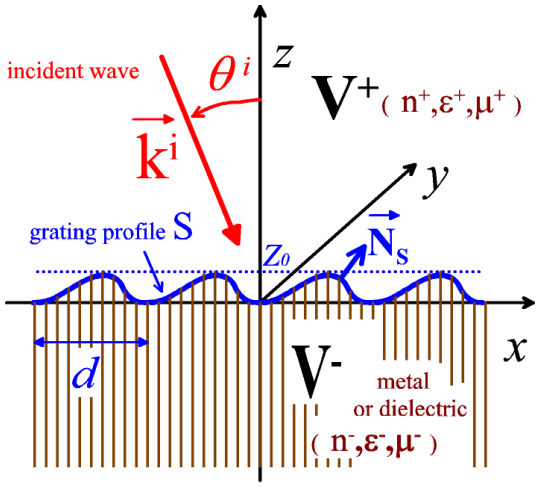

Figure 2.6: Corrugated 1D-periodic grating with period d and analytically described surface S between two homogeneous regions V^+ and V^- [33].

By contrast, the Bloch phase shift can be modelled directly by means of pseudo-periodic boundary conditions in frequency-domain methods. However, even general frequency-domain methods do not directly model the diffractive nature of gratings, i.e. the decomposition Eq. (2.26) of the EM fields in terms of the different diffraction orders.

Due to the high practical importance of diffraction gratings, their wide application and the shortcomings of general numerical methods, several grating-specific algorithms highly suitable for their numerical investigation have been developed. A short overview will be given now and two methods will be highlighted in depth.

A first class of numerical methods for gratings are *integral methods* [33, 34], which are based on surface integral equations deduced from boundary value problems of the theory of scattering. These methods use a surface integral equation of the form

$$v(\mathbf{r}) + \int_S W(\mathbf{r}, \mathbf{r}')u(\mathbf{r}')d\mathbf{r}' = cu(\mathbf{r}) \tag{2.41}$$

to relate the unknown field quantity u at any point in space to its values at the surface of the grating S. Here, v denotes the excitation of the grating, W is the kernel of the equation and the constant $c = 0$ or $c = 1$ determines whether Eq. (2.41) is a Fredholm equation of the first or second kind. These methods are well suited for bare metallic and dielectric, corrugated gratings as per Fig. 2.6, and multilayer gratings with and without interpenetrating layers but they are not directly applicable to gratings with non-analytical function description, like sharp edges or inverted grooves or graded gratings, i.e. gratings where ϵ varies continuously in space.

Another class of methods are *coordinate transform methods*, most notably the Chandezon method (C-method, [35]). The C-method employs a coordinate trans-

form from the Cartesian coordinate system $Oxyz$ to a translation coordinate system $\left(O\bar{r}^1 = x, \bar{r}^2 = y, \bar{r}^3 = z - a(\bar{r}^1)\right)$ where $a(\bar{r}^1)$ is a continuously differentiable function describing the surface profile S of a corrugated grating, see Fig. 2.6. The $\bar{r}^3$ coordinate of the new coordinate system is aligned to the surface of the grating, which greatly simplifies fulfilling the boundary conditions at the grating surface $\bar{r}^3 = 0$. As mentioned in Section 2.2, ME in a curvilinear coordinate system can be described by their covariant form Eq. (2.19). The next steps of the C-method are expressing the EM fields in Eq. (2.19) by means of a pseudo-periodic series as postulated by Bloch's theorem and numerically solving the resulting eigenvalue problem to find the modes of the grating structure. Finally the field amplitudes are matched according to the boundary conditions to obtain the solution of ME. The C-method works well for corrugated structures but can not deal with lamellar or inverted groove gratings.

The *differential theory* [36] of periodic structures uses a projection of the EM field and the structure (in terms of ϵ and μ) onto a set of suitable basis functions in order to reduce Maxwell's equations from a set of partial differential equations to a set of ordinary differential equations. The natural basis for periodic structures and their EM solutions are periodic and pseudo-periodic functions, respectively.

Two methods will be presented in more detail in this thesis: the well established rigorous coupled-wave analysis and the more recently introduced generalised source method. Both these frequency-domain methods directly employ the decomposition of the electromagnetic field into a pseudo-periodic series as given in Eq. (2.25).

The *rigorous coupled-wave analysis* [4, 37–40], also know as *Fourier modal method* (FMM), is a widely used method for the numerical investigation of gratings. The rationale of RCWA is to approximate an arbitrary corrugated diffraction grating by thin periodic layers in which the structure is assumed to be homogeneous in the direction of propagation. In each layer, the EM fields are expressed as a pseudo-periodic series Eq. (2.25) and ME are reformulated as an eigenvalue problem, which is solved numerically. The modes of each layer, i.e. eigensolutions in each layer, are connected by matching the boundary conditions at the interfaces between the layers, which yields the coefficients one needs to express the EM field.

The *generalised source method* [41–44] is a different approach to solving the grating problem. Only a short overview of the GSM is given here as its full derivation for both linear and nonlinear gratings will be given in Chapter 7. The generalised source method is a rigorous method to solve Eq. (2.9) with linear media without sources in the grating region. The main idea of the GSM is to decompose the permittivity $\epsilon(\mathbf{r}) = \epsilon_b(\mathbf{r}) + \Delta\epsilon(\mathbf{r})$ into a simple background structure $\epsilon_b(\mathbf{r})$, for which the analytical solution operator $\mathbf{N}_b$ of ME is known for any source term $\mathbf{J}$, and a difference structure $\Delta\epsilon(\mathbf{r}) = \epsilon(\mathbf{r}) - \epsilon_b(\mathbf{r})$. The

difference permittivity gives rise to a generalised source term $\mathbf{J} = -i\omega\Delta\epsilon(\mathbf{r})\mathbf{E}$ which is defined in terms of the solution $\mathbf{E}$ itself. This yields an implicit description of the solution $\mathbf{E}$:

$$\mathbf{E} = \mathbf{N}_b\left(-i\omega\Delta\epsilon\mathbf{E}\right)$$

This relation is, in the case of the GSM, a system of integral equation. It is discretised by a truncated pseudo-periodic series in the transverse direction and a grid in the longitudinal direction. The resulting linear system of equations is typically large and has to be solved numerically: to obtain a converged solution for a 2D-periodic grating, typically more than 10^5 unknowns are necessary.

Methods for gratings employing coupled integral equations have already been proposed in Refs. [45, 46]. The specific features of the GSM are: *i*) it ensures the correct Fourier factorization which guarantees fast convergence with respect to the number of included diffraction orders in Eq. (2.25) and *ii*) the GSM makes use of the intrinsic structural properties of the linear system of equation to solve it highly efficiently by means of an iterative, matrix-vector multiplication based linear solver.

2.6 Overview of the applications of diffraction gratings and nonlinear optical effects

This section gives an overview of the applications of diffraction gratings and nonlinear optical processes in these photonic devices.

2.6.1 Applications of diffraction gratings

Gratings are used in a multitude of applications, which are directly related to their physical properties, namely the decomposition of the EM fields into the diffraction orders, which enables a convenient way to engineer the optical far-field.

The most common application for gratings as a dispersive optical component stems directly from the decomposition of reflected and transmitted fields into the single diffraction orders in Eq. (2.26). Apart from the central (also called zero or specular) diffraction order, the wavevector of a diffraction order changes with frequency, ω, due to relation Eq. (2.27). Therefore, gratings are widely used as wavelength separation devices in analytical instruments for *spectroscopy*, the study of the interaction between matter and radiated energy. Optical spectroscopy is an invaluable tool in astronomy, analytical chemistry, biology and medicine [47] where grating-based spectrometers are used for qualitative and quantitative analysis. In analytic chemistry, grating based spectrometers can be used to identify atoms and molecules by their inherent discrete spectral absorption lines. They are used in fluorescence spectroscopy and colorimetry. In astronomy, the radiation from distant cosmic

objects can be analysed to gain knowledge of their respective composition, position and velocity. Amongst the main challenges is the low intensity of cosmic signals. Echelle gratings, which posses a triangular profile with high groove angle, are used under grazing incidence to spectrally separate the incoming light in large scale astronomical telescopes [48].

Although diffraction gratings have some limitations for spectroscopy, too, they have several advantages over the classical dispersive element, the prism [33]: The dispersive properties of a prism are defined by the material dispersion of the optical medium, whereas the dispersion in gratings is due to geometric properties. Therefore, gratings can work in spectral regions where there is no transparent and sufficiently dispersive optical material. Moreover, prisms are bulky devices consisting of large volumes of an optically pure material. In contrast, gratings can be plain devices with much smaller size and weight which do not suffer from thermal expansion as strongly and are more suitable for miniaturising the implementation of desired functionalities.

The decomposition of the incident beam into discrete diffraction orders can be used to obtain the functionality of a *multiplexer* or *demultiplexer*, if the neighbouring diffraction orders are viewed as the incident signal. Therefore, diffraction gratings can be used as parallel multiplexing/demultiplexing components in multichannel fibre-optic telecommunication network systems.

While spectroscopy makes use of the noncentral diffraction orders due to the necessary dispersion of their wavevectors, *optical filters* work with the field given by the central diffraction order. The key mechanism is that, depending on the shape and material of the grating structure, incident angle and polarization, more or less power is converted from the incident central order to neighbouring orders or is absorbed in lossy materials. This converted power is not available in the central orders of the reflected or transmitted field. Suitably chosen diffraction gratings made of non-absorbing material can act as broad-band colour filters when operated in transmissions mode [49].

An important application of subwavelength diffraction gratings, that is $\Lambda < \lambda$, is the *grating-assisted coupling* of light to the modes of cavities or waveguides [50, 51]. This is illustrated by the case of a symmetric GaAs slab waveguide of height $h = 500\,\mathrm{nm}$, as per Fig. 2.7(a). The propagation constant β_ν of a slab waveguide mode of order ν is determined by the dispersion relation

$$\kappa(\beta_\nu)h - \arctan\frac{\gamma_2(\beta_\nu)}{\kappa(\beta_\nu)} - \arctan\frac{\gamma_3(\beta_\nu)}{\kappa(\beta_\nu)} = \nu\pi \tag{2.42}$$

with the parameter $\kappa(\beta_\nu) = \sqrt{n_1^2 k_0^2 - \beta_\nu^2}$ and $\gamma_{2,3}(\beta_\nu) = \sigma\sqrt{\beta_\nu^2 - n_{2,3}^2 k_0^2}$ for TE ($\sigma = 1$) and TM ($\sigma = n_1^2/n_2^2$) modes. Here, n_i denotes the index of refraction in the cover ($i = 2$) or substrate regions ($i = 3$) or in the slab waveguide ($i = 1$), and $k_0 = \frac{\omega}{c}$ is the free space

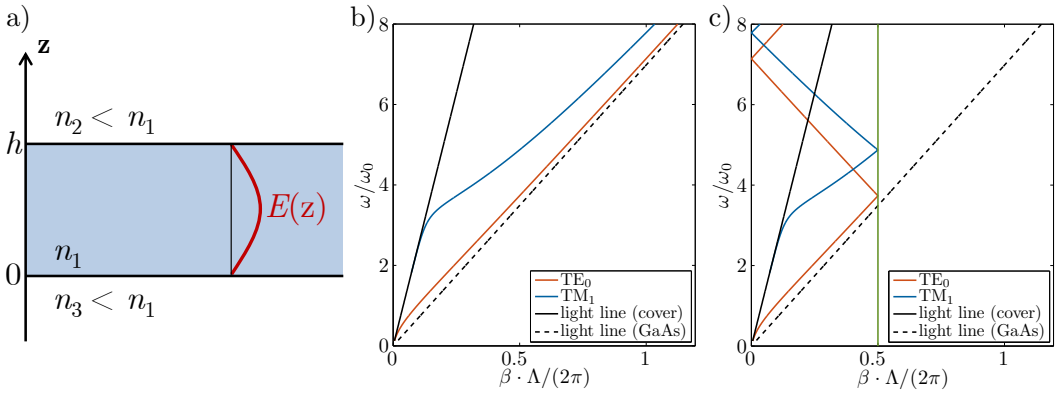

Figure 2.7: a) Schematic representation of an asymmetric slab waveguide and the profile of a mode. b) Dispersion relation of the TE$_0$- TM$_1$-modes of a symmetric GaAs-vacuum slab waveguide and the light line in air. The normalizing angular frequency ω_0 is given by $\omega_0 = c/1\,\mu\text{m}$. c) Dispersion relation folded into the first Brillouin zone.

wavenumber. In order for an incident plane wave to be coupled to a guiding waveguide mode, its parallel wavevector component $k_\parallel$ has to match the propagation constant β_ν of a mode. The dispersion diagram Fig. 2.7(b) shows that this is not possible because the modal propagation constants of the slab waveguide can only be found below the light line of the cover and substrate media. By placing a grating with period $\Lambda = 250\,\text{nm}$ on top of the waveguide, the translational symmetry of the structure is broken and the incident plane wave can be coupled into the waveguide mode whenever the folded dispersion curve Fig. 2.7(c) coincides with the parallel wavevector component:

$$\text{mod}\left(\beta_\nu + k_\parallel + \frac{K}{2}, K\right) = \frac{K}{2}. \tag{2.43}$$

A very detailed treatment of classical diffraction gratings, different types, their analytic properties and further applications, such as tuning the lasing wavelength in laser systems, for pulse compression or broadening, or as beam splitters, can be found in Refs. [47, 52] and the introductory chapters of Ref. [33].

The study and design of *metamaterials* is one of the most thriving topics in modern physics, as metamaterials allow the practical realisation of concepts that were thought to be only possible in theory, as they assume material properties not existing or not yet found in Nature. Due to their novelty, their high potential and the vibrant and ongoing research, it is difficult to formulate a clear definition of metamaterials which is wide enough to cover all aspects while being narrow enough to actually convey a clear meaning [53]. The following definition however serves in many ways: A metamaterial is build up by artificial, engineered structures in the same way natural materials are built up from atoms or molecules. The bulk properties of a natural material are determined by the respective intrinsic and fixed resonant properties of its constituent atoms or molecules. The building blocks in metamaterials are

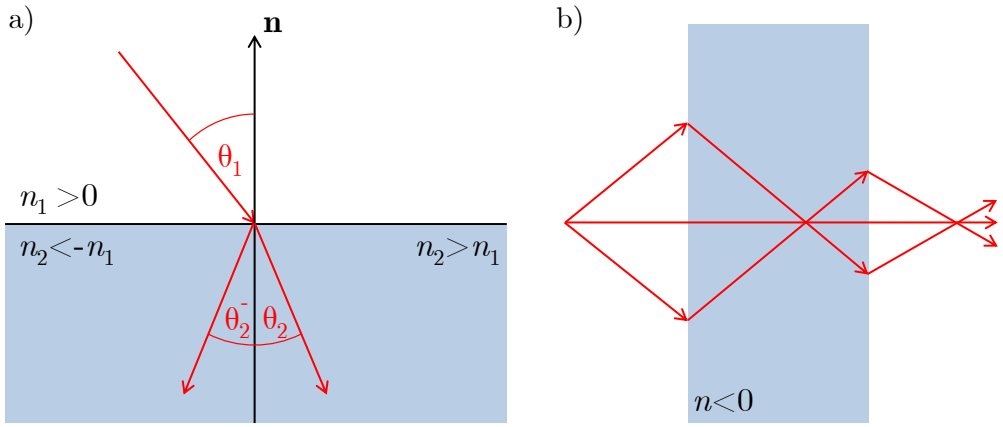

Figure 2.8: a) Refraction of an incident ray at an interface between two positive index materials (right ray) and refraction at an interface between a positive index and a negative index material (left ray): the transmitted ray is on the same side of the surface normal **n** as the incident ray. b) Light formerly diverging from a point source is set in reverse and converges back to a point. Released from the medium the light reaches a focus for a second time (Reproduced from Ref. [55]).

artificial structures, so called meta-atoms or meta-molecules, which can be engineered to achieve nearly any macroscopic property of the metamaterial. If a metamaterial is to be interpreted as an effective medium, the size of its meta-atoms has to be significantly smaller than the operating wavelength.

The most compelling example for the potential of metamaterials is the negative index material which was first theoretically envisioned by Veselago [54] in 1968. It is defined as a medium with both $\epsilon_r < 0$ and $\mu_r < 0$ resulting in a negative refractive index $n = -\sqrt{\epsilon_r \mu_r}$. This assumption has interesting implications on the propagation of light in such materials. For example, the transmission of light at the interface between a conventional material $n_1 > 0$ and a negative index material $n_2 < 0$ is still described by the conventional Snell law

$$\frac{\sin\theta_1}{\sin\theta_2} = \frac{n_2}{n_1}. \tag{2.44}$$

Light at such an interface is refracted in the opposite direction (see Fig. 2.8(a) left ray) as to what is expected for regular materials (Fig. 2.8(a) right ray).

An important direct application of such a negative index material is subwavelength imaging, i.e. the construction of superlenses with infinitely fine resolution [55], see Fig. 2.8(b) which surpass Abbe's diffraction limit. However, there is no naturally occurring negative-index material and it took until 2001 to design a metamaterial to effectively behave as a negative index material, as per Fig. 2.9, which shows the structure used in the first experimental demonstration of negative index of refraction in the microwave wavelength range [56].

Apart from negative index materials, properly engineered metamaterials could be used

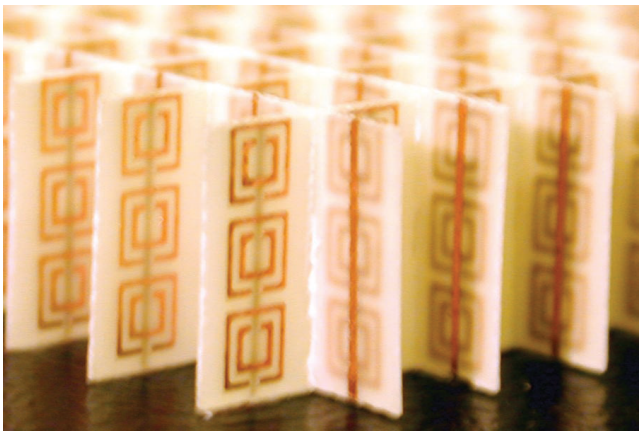

Figure 2.9: Photo of a negative-index metamaterial [56] consisting of a lattice of square copper split ring resonators and copper wire strips on opposite sides of fibre glass circuit boards. This structure effectively acts as a negative index material at microwave frequencies.

to realise some exotic material properties demanded by transformation optics [7], find use in highly efficient antennas, cloaking devices, or as highly efficient absorbers in photovoltaic applications.

Although there are also non-periodic metamaterial designs (e.g. with a Penrose tiling) the above negative index material example and many other designs of metamaterials are periodic. Therefore, the study of metamaterials is inherently linked with that of diffraction gratings and many *diffraction gratings* can be seen as metamaterials or more exactly *metasurfaces*.

2.6.2 Selected applications of nonlinear optical phenomena

In the regime of linear optics, the propagation of light with one frequency is independent of its intensity and the distribution of light with other frequencies and the total optical field is given by linear superposition of the individual frequency contributions. This principle of linear superposition does not hold in the nonlinear optical regime due to the different nonlinear phenomena described in Section 2.4. These nonlinear optical effects could hence be seen purely as an added complexity to the simpler, better understood theory of linear optics. However, just these nonlinear mechanisms open the way for a multitude of applications, which are impossible to achieve by means of linear optics, and several examples will be discussed in this section.

First and foremost is the possibility for *frequency conversion*, a functionality that is impossible to achieve using linear optical processes. This is illustrated by a neodymium-doped yttrium aluminium garnet (Nd:YAG) solid-state laser [57]. The direct output of a Nd:YAG laser at $1064\,\mathrm{nm}$ can be converted to $532\,\mathrm{nm}$ by frequency locking the laser to a monolithic ring resonator constructed from magnesium oxide-doped lithium niobate, MgO:$LiNbO_3$, which is a highly nonlinear optical crystal. This system yields an excellent conversion

efficiency of 65% from fundamental to second-harmonic with no signs of temporal degradation, making this combined Nd:YAG-MgO:$LiNbO_3$ laser system an excellent choice for generating laser light at a frequency in the visible range.

While most of the early studies of higher-harmonic generation focused on wave interaction in homogeneous nonlinear optical media, recent advances in nanofabrication techniques and the advent of metamaterials have considerably broadened the experimental and theoretical framework in which these nonlinear optical processes are explored. In particular, second-harmonic generation from isotropic [58–60] and chiral [61, 62] metasurfaces, layered media coupled to arrays of plasmonic particles [63], and quantum engineered plasmonic metasurfaces [64] have been investigated.

As seen in Section 2.4, nonlinear optical processes require high intensities to occur. The diffractive properties of gratings can be used to achieve local field enhancement and hence improve the efficiency of nonlinear application.

Second- and third-harmonic generation are widely used in nonlinear *optical microscopy* [15]. One advantage of employing nonlinear mechanisms is a high resolution due to the localization of the harmonic generation to regions with maximal electric field intensity of a focused laser beam. Another feature is that any unwanted influence of the pump beam can easily be removed from the measured signal, because signal and pump are spectrally separated. Nonlinear optical microscopy is intensively utilized in biomedical applications where either the intrinsic nonlinear properties of the biological sample materials or those of artificially induced marker particles are employed [65]. These techniques have been used for non-destructive, non-invasive imaging of living tissue cells both ex- and in vivo and helped to identify numerous pathologies e.g. related to collagen disorder in skin [66] or cancerous diseases [66, 67].

Nonlinear optical effects play an important role in *fibre optics* [68]. Glass as a centrosymmetric material only allows nonlocal SHG and surface SHG, see Ref. [17]. Hence, the most important nonlinear effects stem from third-order processes: nonlinear refraction Eq. (2.36), which has to be accounted for in the design of fibre systems, and four-wave mixing (FWM). They influences fibre optical systems in several ways. FWM as a feature is used for fibre-based optical parametric amplification, for parametric oscillators and wavelength conversion. However, FWM and especially nonlinear refraction degrade the efficiency of wavelength-division multiplexing and are a cause for interchannel crosstalk in multichannel optical fibre communication systems, which generally should be avoided.

The angular distribution and intensity of *surface nonlinear effects* depends heavily on the morphology of the surface and is highly sensitive to the presence of imperfections or impurities. Hence, the principles of surface nonlinear optics are the basis for highly selective and sensitive techniques to optically *probe interfacial phenomena*. The enhancement of

efficiency of SHG within extremely small nanoscopic volumes is of paramount interest in surface science [17, 69], colloidal chemistry [70], and catalytic chemistry [71]. Due to its nature, surface second harmonic generation is extremely sensitive to surface adsorbents. As such, even a single-molecule layer adsorbed onto a surface can completely change the surface nonlinear susceptibility. This sensitivity to changes of the chemical environment has found many applications to the study of the symmetry properties of surfaces [72], the nature of adsorbates at surfaces or interfaces [73], or noninvasive probing of buried interfaces [69].

Optical activity is a phenomenon that manifests itself in two ways: optical rotation, i.e. the change of polarization angle of a plane wave during its propagation through a medium, and circular dichroism (CD), i.e. different interaction of left- and right-circular polarized light with a medium. These phenomena occur in chiral media. An object (e.g. a molecule or a nanostructure) is chiral if it can not be superimposed on its mirror image by rotations and translations. Therefore, optical activity is intimately related to the symmetry properties of chemical substances, biomaterials and crystals and has been successfully employed to probe them [74, 75]. Moreover, it can be instrumental in assessing the chiral purity of pharmaceuticals, agrochemicals and biomolecules [76–79]. Typically, the optical activity of chiral media is very small, and ways to enhance it are necessary. Plasmonic field enhancement at the surface of metal nanostructures [76–78] can increase the interaction of light with chiral molecules. The optical activity for higher-harmonic generation exhibits much higher contrast for the two circular polarizations [79–82] as compared to the linear optical activity and offers hence a powerful technique to probe the symmetry of organic and inorganic materials due to its high sensitivity to the asymmetry of molecules and physical interfaces [83–85].

2.7 Introduction to 2D materials

The physical and theoretical concepts that have been introduced so far and their applications have been concerned with conventional, three dimensional media, i.e. media that have a spatial extent in all three spatial dimensions. This section will introduce the novel class of two dimensional, atomically thin materials, which exhibit unique and intriguing novel physical properties and have hence entered the focus of current research.

Since its first isolation, preparation, and theoretical description [86–88], graphene, a monolayer of carbon atoms distributed in a hexagonal lattice, has attracted a tremendous amount of interest in science and engineering due primarily to its outstanding physical properties and potential for novel or optimised applications. Graphene was shown to have remarkable mechanic strength [89–91] and extremely high thermal conductivity [92], making it a particularly appealing materials platform for nano-electromechanical applications and management of thermal processes in nano-electronic circuits [90, 93]. In addition, the high carrier mobility of graphene enhances its potential for applications to high-frequency electronics [94–96]. These and other remarkable properties of an atomically thin mono-

layer of graphite have spurred new research into and development of new two-dimensional materials, such as hexagonal boron nitride (*h*-BN), silicene (monolayers of silicon), and transition metal dichalcogenide (TMDC) monolayers [97–100], each with their own series of unique physical properties.

One additional compelling aspect of 2D materials is closely related to their optical properties. Graphene, for example, is nearly transparent at optical frequencies, exhibiting absorption of 2.3% of white light [101], which suggests it barely interacts with light. This transparency and the earlier mentioned electro-mechanical properties make graphene a promising new material for flexible optical devices [102] (e.g. touch screens). Moreover, graphene based structures can serve as an alternative to conventional metallo-dielectric structures to spatially confine and guide light in the emerging field of graphene nanoplasmonics [103–111].

In addition to these linear properties, the nonlinear optical properties of graphene and other 2D materials have attracted increased attention. Graphene, as a centrosymmetric material, exhibits third-harmonic generation [112, 113], strong optical Kerr nonlinearity [114], and induced second order nonlinearity [115–117] in a single atomic layer if placed on a substrate. This allows applications of graphene as ultra-compact optical modulator, optical limiter, as frequency converter and in photovoltaic and photoresistive devices [102, 112–115, 118]. On the other hand, other 2D materials, e.g semiconducting TMDC monolayers, are non-centrosymmetric and hence allow even-order nonlinear optical processes [119–122]. Due to their optical and electrical properties, they are suggested as nanoscale transistors and saturable absorbers [123, 124]. The implementation of these linear and nonlinear optical properties into applications, however, requires advances of fabrication and experimental techniques [102, 125, 126], theoretical models, and, finally, computational methods for modelling of devices incorporating 2D materials.

Bibliography

[1] J. C. Maxwell, "A Dynamical Theory of the Electromagnetic Field," Proceedings of the Royal Society of London **13**, 531–536 (1863).

[2] J. D. Jackson, *Classical Electrodynamics*, 3rd ed. (John Wiley, 1999).

[3] P. Halevi, *Spatial Dispersion in Solids and Plasmas*, Electromagnetic waves (North-Holland, 1992).

[4] L. Li, "Fourier modal method for crossed anisotropic gratings with arbitrary permittivity and permeability tensors," J. Opt. Soc. Am. A. **5**, 345 (2003).

[5] S. Essig and K. Busch, "Generation of adaptive coordinates and their use in the fourier modal method," Opt. Express **18**, 23258–23274 (2010).

[6] M. I. Mishchenko, L. D. Travis, and D. W. Mackowski, "T-matrix computations of light scattering by nonspherical particles: a review," J. Quant. Spectrosc. Radiat. Transfer **55**, 535–575 (1996).

[7] H. Chen, C. Chan, and P. Sheng, "Transformation optics and metamaterials," Nat. Mater. **9**, 387–396 (2010).

[8] J. Joannopoulos, S. Johnson, J. Winn, and R. Meade, *Photonic Crystals: Molding the Flow of Light*, 2nd ed. (Princeton University Press, 2011).

[9] K. Sakoda, *Optical Properties of Photonic Crystals*, 2nd ed., Springer Series in Optical Sciences (Springer, 2004).

[10] J. Singleton, *Band theory and electronic properties of solids*, Oxford master series in condensed matter physics (Oxford University Press, 2001).

[11] C. Kittel, *Introduction to Solid State Physics*, 8th ed. (Wiley, 2004).

[12] R. W. Wood, "On a remarkable case of uneven distribution of light in a diffraction grating spectrum," Proceedings of the Physical Society of London **18**, 269 (1902).

[13] P. Franken, A. Hill, C. Peters, and G Weinreich, "Generation of optical harmonics," Phys. Rev. Lett. **7**, 118–119 (1961).

[14] S. Bulanov, T. Esirkepov, and T. Tajima, "Light intensification towards the schwinger limit," Phys. Rev. Lett. **91**, 085001 (2003).

[15] R. W. Boyd, *Nonlinear optics* (Academic Press, 2008).

[16] Y. Shen, *The principles of nonlinear optics*, Wiley classics library (Wiley-Interscience, 2003).

[17] T. Heinz, "Nonlinear surface electromagnetic phenomena," in (Elsevier, Oxford, 1991) Chap. Second-Order Nonlinear Optical Effects at Surfaces and Interfaces.

[18] N. Bloembergen and P. Pershan, "Light waves at the boundary of nonlinear media," Phys. Rev. **128**, 606–622 (1962).

[19] P. Mouroulis and J. Macdonald, *Geometrical optics and optical design*, Vol. 39 (Oxford University Press New York, 1997).

[20] M. Born and E. Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light* (Elsevier Science Limited, 1980), p. 808.

[21] W. C. Gibson, *The method of moments in electromagnetics* (CRC Press, 2007).

[22] B. T. Draine and P. J. Flatau, "Discrete-dipole approximation for scattering calculations," J. Opt. Soc. Am. A. **11**, 1491–1499 (1994).

[23] K. Yee, "Numerical solution of inital boundary value problems involving Maxwell's equations in isotropic media," IEEE Trans. Antennas Propag. **14**, 302–307 (1966).

[24] E. Haber and U. M. Ascher, "Fast finite volume simulation of 3d electromagnetic problems with highly discontinuous coefficients," SIAM J. Sci. Comput. **22**, 1943–1961 (2001).

[25] A. Taflove and S. C. Hagness, *Computational electrodynamics* (Artech House, 2005).

[26] Lumerical Solutions, Inc., *FDTD Solutions*, `http://www.lumerical.com/tcad-products/fdtd/`.

[27] Photon Design Ltd., *OmniSim/FDTD*, `http://www.photond.com/products/omnisim.htm`.

[28] A. F. Oskooi, D. Roundy, M. Ibanescu, P. Bermel, J. D. Joannopoulos, and S. G. Johnson, "MEEP: A flexible free-software package for electromagnetic simulations by the FDTD method," Comput. Phys. Commun. **181**, 687–702 (2010).

[29] G. Sztefka and H. Nolting, "Bidirectional eigenmode propagation for large refractive index steps," IEEE Photon. Technol. Lett. **5**, 554–557 (1993).

[30] O. V. Sinkin, R. Holzlöhner, J. Zweck, and C. R. Menyuk, "Optimization of the split-step fourier method in modeling optical-fiber communications systems," IEEE J. Lightwave Technol. **21**, 61–68 (2003).

[31] P. A. Martin, *Multiple scattering: interaction of time-harmonic waves with n obstacles* (Cambridge University Press, 2006).

[32] D. Felbacq, G. Tayeb, and D. Maystre, "Scattering by a random set of parallel cylinders," J. Opt. Soc. Am. A. **11**, 2526–2538 (1994).

[33] E. Popov, ed., *Gratings: theory and numeric applications* (Popov, Institut Fresnel, 2012).

[34] D. Maystre, "A new general integral theory for dielectric coated gratings," J. Opt. Soc. Am. **68**, 490–495 (1978).

[35] J. Chandezon, M. Dupuis, G. Cornet, and D. Maystre, "Multicoated gratings: a differential formalism applicable in the entire optical region," J. Opt. Soc. Am. **72**, 839–846 (1982).

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

[37] M. G. Moharam, T. K. Gaylord, 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).

[38] L. Li, "Use of fourier series in the analysis of discontinuous periodic structures," J. Opt. Soc. Am. A. **13**, 1870–1876 (1996).

[39] L. Li, "New formulation of the fourier modal method for crossed surface-relief gratings," J. Opt. Soc. Am. A. **14**, 2758–2767 (1997).

[40] M. Weismann, D. F. G. Gallagher, and N. C. Panoiu, "Accurate near-field evaluation in the rigorous coupled-wave analysis," J. Opt. **17**, 125612 (2015).

[41] A. A. Shcherbakov and A. V. Tishchenko, "Fast numerical method for modelling one-dimensional diffraction gratings," IEEE J. Quantum Electron. **40**, 538 (2010).

[42] A. A. Shcherbakov and A. V. Tishchenko, "New fast and memory-sparing method for rigorous electromagnetic analysis of 2d periodic dielectric structures," J. Quant. Spectrosc. Radiat. Transfer **113**, 158–171 (2012).

[43] A. A. Shcherbakov and A. V. Tishchenko, "Efficient curvilinear coordinate method for grating diffraction simulation," Opt. Express **21**, 25236–25247 (2013).

[44] M. Weismann, D. F. G. Gallagher, and N. C. Panoiu, "Nonlinear generalized source method for modeling second-harmonic generation in diffraction gratings," J. Opt. Soc. Am. B. **32**, 523–533 (2015).

[45] A. Cowan and J. F. Young, "Mode matching for second-harmonic generation in photonic crystal waveguides," Phys. Rev. B **65**, 085106 (2002).

[46] T. Magath and A. E. Serebryannikov, "Fast iterative, coupled-integral-equation technique for inhomogeneous profiled and periodic slabs," J. Opt. Soc. Am. A. **22**, 2405–2418 (2005).

[47] C. A. Palmer and E. G. Loewen, eds., *Diffraction grating handbook* (Newport Corporation, Springfield, Ohio, 2005).

[48] S. S. Vogt, S. L. Allen, B. C. Bigelow, L. Bresee, B. Brown, T. Cantrall, A. Conrad, M. Couture, C. Delaney, H. W. Epps, D. Hilyard, D. F. Hilyard, E. Horn, N. Jern, D. Kanto, M. J. Keane, R. I. Kibrick, J. W. Lewis, J. Osborne, G. H. Pardeilhan, T. Pfister, T. Ricketts, L. B. Robinson, R. J. Stover, D. Tucker, J. Ward, and M. Z. Wei, "HIRES: the high-resolution echelle spectrometer on the Keck 10-m Telescope," in 1994 symposium on astronomical telescopes & instrumentation for the 21st century (International Society for Optics and Photonics, 1994), pp. 362–375.

[49] K. Knop, "Diffraction gratings for color filtering in the zero diffraction order," Appl. Opt. **17**, 3598–3603 (1978).

[50] R. R. A. Syms and J. R. Cozens, *Optical guided waves and devices* (McGraw-Hill, 1992).

[51] N. C. Panoiu and R. M. Osgood, "Enhanced optical absorption for photovoltaics via excitation of waveguide and plasmon-polariton modes," Opt. Lett. **32**, 2825–2827 (2007).

[52] E. Loewen and E. Popov, *Diffraction gratings and applications*, Optical Science and Engineering (Taylor & Francis, 1997).

[53] T. J. Cui, D. Smith, and R. Liu, *Metamaterials: theory, design, and applications* (Springer, 2009).

[54] V. G. Veselago, "The electrodynamics of substances with simultaneously negative values of and ε and μ," Sov. Phys. Usp. **10**, 509 (1968).

[55] J. B. Pendry, "Negative refraction makes a perfect lens," Phys. Rev. Lett. **85**, 3966 (2000).

[56] R. A. Shelby, D. R. Smith, and S. Schultz, "Experimental verification of a negative index of refraction," Science **292**, 77–79 (2001).

[57] D. C. Gerstenberger, G. E. Tye, and R. W. Wallace, "Efficient second-harmonic conversion of cw single-frequency nd:yag laser light by frequency locking to a monolithic ring frequency doubler," Opt. Lett. **16**, 992–994 (1991).

[58] W. Fan, S. Zhang, N.-C. Panoiu, A Abdenour, S Krishna, R. Osgood, K. Malloy, and S. Brueck, "Second harmonic generation from a nanopatterned isotropic nonlinear material," Nano Lett. **6**, 1027–1030 (2006).

[59] M. W. Klein, C. Enkrich, M. Wegener, and S. Linden, "Second-harmonic generation from magnetic metamaterials," Science **313**, 502–504 (2006).

[60] J. A. H. Van Nieuwstadt, M Sandtke, R. H. Harmsen, F. B. Segerink, J. C. Prangsma, S. Enoch, and L Kuipers, "Strong modification of the nonlinear optical response of metallic subwavelength hole arrays," Phys. Rev. Lett. **97**, 146102 (2006).

[61] V. K. Valev, A. V. Silhanek, N. Verellen, W. Gillijns, P. Van Dorpe, O. A. Aktsipetrov, G. A. E. Vandenbosch, V. V. Moshchalkov, and T. Verbiest, "Asymmetric Optical Second-Harmonic Generation from Chiral G-Shaped Gold Nanostructures," Phys. Rev. Lett. **104**, 127401 (2010).

[62] V. K. Valev, J. J. Baumberg, B. De Clercq, N. Braz, X. Zheng, E. J. Osley, S. Vandendriessche, M. Hojeij, C. Blejean, J. Mertens, C. G. Biris, V. Volskiy, M. Ameloot, Y. Ekinci, G. A. E. Vandenbosch, P. A. Warburton, V. V. Moshchalkov, N. C. Panoiu, and T. Verbiest, "Nonlinear superchiral meta-surfaces: tuning chirality and disentangling non-reciprocity at the nanoscale," Adv. Mater. **26**, 4074–4081 (2014).

[63] K. Chen, C. Durak, J. Heflin, and H. D. Robinson, "Plasmon-enhanced second-harmonic generation from ionic self-assembled multilayer films," Nano Lett. **7**, 254–258 (2007).

[64] J. Lee, M. Tymchenko, C. Argyropoulos, P.-Y. Chen, F. Lu, F. Demmerle, G. Boehm, M.-C. Amann, A. Alù, and M. A. Belkin, "Giant nonlinear response from plasmonic metasurfaces coupled to intersubband transitions," Nature **511**, 65–69 (2014).

[65] F. Pavone and P. Campagnola, *Second harmonic generation imaging*, Series in Cellular and Clinical Imaging (Taylor & Francis, 2013).

[66] R. M. Williams, W. R. Zipfel, and W. W. Webb, "Interpreting second-harmonic generation images of collagen i fibrils," Biophys. J. **88**, 1377–1386 (2005).

[67] E. Brown, T. McKee, A. Pluen, B. Seed, Y. Boucher, and R. K. Jain, "Dynamic imaging of collagen and its modulation in tumors in vivo using second-harmonic generation," Nat. Med. **9**, 796–800 (2003).

[68] G. Agrawal, *Nonlinear fiber optics*, Optics and Photonics (Elsevier Science, 2001).

[69] K. Eisenthal, "Liquid interfaces probed by second-harmonic and sum-frequency spectroscopy," Chem. Rev **96**, 1343–1360 (1996).

[70] K. B. Eisenthal, "Second harmonic spectroscopy of aqueous nano- and microparticle interfaces," Chemical Reviews **106**, 1462–1477 (2006).

[71] R. M. Corn and D. A. Higgins, "Optical second harmonic generation as a probe of surface chemistry," Chem. Rev. **94**, 107–125 (1994).

[72] T. F. Heinz, M. M. T. Loy, and W. A. Thompson, "Study of si(111) surfaces by optical second-harmonic generation: reconstruction and surface phase transformation," Phys. Rev. Lett. **54**, 63–66 (1985).

[73] H. W. K. Tom, C. M. Mate, X. D. Zhu, J. E. Crowell, T. F. Heinz, G. A. Somorjai, and Y. R. Shen, "Surface Studies by Optical Second-Harmonic Generation: The Adsorption of O_2, CO, and Sodium on the Rh(111) Surface," Phys. Rev. Lett. **52**, 348–351 (1984).

[74] R. M. Hazen and D. S. Sholl, "Chiral selection on inorganic crystalline surfaces," Nat. Mater. **2**, 367–374 (2003).

[75] K.-H. Ernst, "Molecular chirality at surfaces," Phys. Status Solidi B **249**, 2057–2088 (2012).

[76] B. M. Maoz, Y. Chaikin, A. B. Tesler, O. Bar Elli, Z. Fan, A. O. Govorov, and G. Markovich, "Amplification of Chiroptical Activity of Chiral Biomolecules by Surface Plasmons," Nano Lett. **13**, 1203–1209 (2013).

[77] R. Schreiber, N. Luong, Z. Fan, A. Kuzyk, P. C. Nickels, T. Zhang, D. M. Smith, B. Yurke, W. Kuang, A. O. Govorov, and T. Liedl, "Chiral plasmonic DNA nanostructures with switchable circular dichroism," Nat. Commun. **4**, 2948 (2013).

[78] F. Lu, Y. Tian, M. Liu, D. Su, H. Zhang, A. O. Govorov, and O. Gang, "Discrete Nanocubes as Plasmonic Reporters of Molecular Chirality," Nano Lett. **13**, 3145–3151 (2013).

[79] S. Chen, F. Zeuner, M. Weismann, B. Reineke, G. Li, V. K. Valev, K. W. Cheah, N. C. Panoiu, T. Zentgraf, and S. Zhang, "Giant Nonlinear Optical Activity of Achiral Origin in Planar Metasurfaces with Quadratic and Cubic Nonlinearities," Adv. Mater. (2016).

[80] T. Verbiest, M. Kauranen, and A. Persoons, "Light-Polarization-Induced Optical Activity," Phys. Rev. Lett. **82**, 3601–3604 (1999).

[81] T Petralli-Mallow, T. Wong, J. Byers, H. Yee, and J. Hicks, "Circular dichroism spectroscopy at interfaces: a surface second harmonic generation study," J. Phys. Chem. **97**, 1383–1388 (1993).

[82] J. D. Byers, H. I. Yee, and J. M. Hicks, "A second harmonic generation analog of optical rotatory dispersion for the study of chiral monolayers," J. Chem. Phys. **101**, 6233–6241 (1994).

[83] J. J. Maki, M. Kauranen, and A. Persoons, "Surface second-harmonic generation from chiral materials," Phys. Rev. B **51**, 1425–1434 (1995).

[84] T. Verbiest, M. Kauranen, Y. van Rompaey, and A. Persoons, "Optical Activity of Anisotropic Achiral Surfaces," Phys. Rev. Lett. **77**, 1456–1459 (1996).

[85] P. Fischer, "Comprehensive chiroptical spectroscopy," in, Vol. 1: Instrumentation, Methodologies, and Theoretical Simulations ((Eds: N. Berova, P. L. Polavarapu, K. Nakanishi, R. W. Woody), John Wiley & Sons, Inc., Hoboken, NJ, USA, 2012).

[86] K. S. Novoselov, A. K. Geim, S. Morozov, D Jiang, Y Zhang, S. Dubonos, I. Grigorieva, and A. Firsov, "Electric field effect in atomically thin carbon films," Science **306**, 666–669 (2004).

[87] K. Novoselov, A. K. Geim, S. Morozov, D Jiang, M. Katsnelson, I. Grigorieva, S. Dubonos, and A. Firsov, "Two-dimensional gas of massless Dirac fermions in graphene," Nature **438**, 197–200 (2005).

[88] K. S. Novoselov, D. Jiang, F. Schedin, T. J. Booth, V. V. Khotkevich, S. V. Morozov, and A. K. Geim, "Two-dimensional atomic crystals," Proc. Natl. Acad. Sci. USA **102**, 10451–10453 (2005).

[89] C. Lee, X. Wei, J. W. Kysar, and J. Hone, "Measurement of the elastic properties and intrinsic strength of monolayer graphene," Science **321**, 385–388 (2008).

[90] I. Frank, D. M. Tanenbaum, A. Van der Zande, and P. L. McEuen, "Mechanical properties of suspended graphene sheets," J. Vac. Sci. Technol. B **25**, 2558–2561 (2007).

[91] J. C. Meyer, A. K. Geim, M. Katsnelson, K. Novoselov, T. Booth, and S Roth, "The structure of suspended graphene sheets," Nature **446**, 60–63 (2007).

[92] A. A. Balandin, S. Ghosh, W. Bao, I. Calizo, D. Teweldebrhan, F. Miao, and C. N. Lau, "Superior thermal conductivity of single-layer graphene," Nano Lett. **8**, 902–907 (2008).

[93] S Ghosh, I Calizo, D Teweldebrhan, E. Pokatilov, D. Nika, A. Balandin, W Bao, F Miao, and C. N. Lau, "Extremely high thermal conductivity of graphene: Prospects for thermal management applications in nanoelectronic circuits," Appl. Phys. Lett. **92**, 151911 (2008).

[94] S. Morozov, K. Novoselov, M. Katsnelson, F Schedin, D. Elias, J. Jaszczak, and A. Geim, "Giant intrinsic carrier mobilities in graphene and its bilayer," Phys. Rev. Lett. **100**, 016602 (2008).

[95] Y.-M. Lin, C. Dimitrakopoulos, K. A. Jenkins, D. B. Farmer, H.-Y. Chiu, A. Grill, and P. Avouris, "100-GHz transistors from wafer-scale epitaxial graphene," Science **327**, 662–662 (2010).

[96] F. Schwierz, "Graphene transistors," Nature Nanotech. **5**, 487–496 (2010).

[97] M. Xu, T. Liang, M. Shi, and H. Chen, "Graphene-Like Two-Dimensional Materials," Chem. Rev. **113**, 3766–3798 (2013).

[98] L. Song, L. Ci, H. Lu, P. B. Sorokin, C. Jin, J. Ni, A. G. Kvashnin, D. G. Kvashnin, J. Lou, B. I. Yakobson, and P. M. Ajayan, "Large Scale Growth and Characterization of Atomic Hexagonal Boron Nitride Layers," Nano Lett. **10**, 3209–3215 (2010).

[99] P. Vogt, P. De Padova, C. Quaresima, J. Avila, E. Frantzeskakis, M. C. Asensio, A. Resta, B. Ealet, and G. Le Lay, "Silicene: compelling experimental evidence for graphenelike two-dimensional silicon," Phys. Rev. Lett. **108**, 155501 (2012).

[100] Q. H. Wang, K. Kalantar-Zadeh, A. Kis, J. N. Coleman, and M. S. Strano, "Electronics and optoelectronics of two-dimensional transition metal dichalcogenides," Nature Nanotech. **7**, 699–712 (2012).

[101] R. Nair, P. Blake, A. Grigorenko, K. Novoselov, T. Booth, T. Stauber, N. Peres, and A. Geim, "Fine structure constant defines visual transparency of graphene," Science **320**, 1308–1308 (2008).

[102] F. Bonaccorso, Z. Sun, T. Hasan, and A. Ferrari, "Graphene photonics and optoelectronics," Nature Photon. **4**, 611–622 (2010).

[103] D. A. Smirnova, I. V. Shadrivov, A. E. Miroshnichenko, A. I. Smirnov, and Y. S. Kivshar, "Second-harmonic generation by a graphene nanoparticle," Phys. Rev. B **90**, 035412 (2014).

[104] M. Jablan, H. Buljan, and M. Soljačić, "Plasmonics in graphene at infrared frequencies," Phys. Rev. B **80**, 245435 (2009).

[105] L. Ju, B. Geng, J. Horng, C. Girit, M. Martin, Z. Hao, H. A. Bechtel, X. Liang, A. Zettl, Y. R. Shen, et al., "Graphene plasmonics for tunable terahertz metamaterials," Nature Nanotech. **6**, 630–634 (2011).

[106] Q. Bao and K. P. Loh, "Graphene photonics, plasmonics, and broadband optoelectronic devices," ACS nano **6**, 3677–3694 (2012).

[107] M. Jablan, M. Soljai, and H. Buljan, "Plasmons in Graphene: Fundamental Properties and Potential Applications," Proc. IEEE **101**, 1689–1704 (2013).

[108] X. Luo, T. Qiu, W. Lu, and Z. Ni, "Plasmons in graphene: recent progress and applications," Mat Sci Eng R **74**, 351 –376 (2013).

[109] K. Y. Yeung, J. Chee, H. Yoon, Y. Song, J. Kong, and D. Ham, "Far-infrared graphene plasmonic crystals for plasmonic band engineering," Nano Lett. **14**, 2479–2484 (2014).

[110] D. A. Smirnova, A. E. Miroshnichenko, Y. S. Kivshar, and A. B. Khanikaev, "Tunable nonlinear graphene metasurfaces," Phys. Rev. B **92**, 161406 (2015).

[111] D. A. Smirnova, R. E. Noskov, L. A. Smirnov, and Y. S. Kivshar, "Dissipative plasmon solitons in graphene nanodisk arrays," Phys. Rev. B **91**, 075409 (2015).

[112] S.-Y. Hong, J. I. Dadap, N. Petrone, P.-C. Yeh, J. Hone, and R. M. Osgood Jr, "Optical third-harmonic generation in graphene," Phys. Rev. X **3**, 021014 (2013).

[113] J. L. Cheng, N Vermeulen, and J. E. Sipe, "Third order optical nonlinearity of graphene," New J. Phys. **16**, 053014 (2014).

[114] H. Zhang, S. Virally, Q. Bao, L. K. Ping, S. Massar, N. Godbout, and P. Kockaert, "Z-scan measurement of the nonlinear refractive index of graphene," Opt. Lett. **37**, 1856–1858 (2012).

[115] A. Y. Bykov, T. V. Murzina, M. G. Rybin, and E. D. Obraztsova, "Second harmonic generation in multilayer graphene induced by direct electric current," Phys. Rev. B **85**, 121413 (2012).

[116] J. Cheng, N. Vermeulen, and J. Sipe, "DC current induced second order optical nonlinearity in graphene," Opt. Express **22**, 15868–15876 (2014).

[117] Y. Q. An, J. Rowe, D. B. Dougherty, J. U. Lee, and A. C. Diebold, "Optical second-harmonic generation induced by electric current in graphene on Si and SiC substrates," Phys. Rev. B **89**, 115310 (2014).

[118] F. Xia, T. Mueller, Y.-m. Lin, A. Valdes-Garcia, and P. Avouris, "Ultrafast graphene photodetector," Nature Nanotech. **4**, 839–843 (2009).

[119] C. Janisch, Y. Wang, D. Ma, N. Mehta, A. L. Elas, N. Perea-Lpez, M. Terrones, V. Crespi, and Z. Liu, "Extraordinary Second Harmonic Generation in Tungsten Disulfide Monolayers," Sci. Rep. **4**, 5530 (2014).

[120] L. M. Malard, T. V. Alencar, A. P. M. Barboza, K. F. Mak, and A. M. de Paula, "Observation of intense second harmonic generation from MoS_2 atomic crystals," Phys. Rev. B **87**, 201401 (2013).

[121] Y. Li, Y. Rao, K. F. Mak, Y. You, S. Wang, C. R. Dean, and T. F. Heinz, "Probing Symmetry Properties of Few-Layer MoS_2 and h-BN by Optical Second-Harmonic Generation," Nano Lett. **13**, 3329–3333 (2013).

[122] K. L. Seyler, J. R. Schaibley, P. Gong, P. Rivera, A. M. Jones, S. Wu, J. Yan, D. G. Mandrus, W. Yao, and X. Xu, "Electrical control of second-harmonic generation in a WSe_2 monolayer transistor," Nat. Nanotechnol. **10**, 407–411 (2015).

[123] B. Radisavljevic, A. Radenovic, J. Brivio, V. Giacometti, and A. Kis, "Single-layer MoS_2 transistors," Nature Nanotech. **6**, 147–150 (2011).

[124] H. Zhang, S. Lu, J. Zheng, J. Du, S. Wen, D. Tang, and K. Loh, "Molybdenum disulfide (MoS_2) as a broadband saturable absorber for ultra-fast photonics," Opt. Express **22**, 7249–7260 (2014).

[125] K. S. Kim, Y. Zhao, H. Jang, S. Y. Lee, J. M. Kim, K. S. Kim, J.-H. Ahn, P. Kim, J.-Y. Choi, and B. H. Hong, "Large-scale pattern growth of graphene films for stretchable transparent electrodes," Nature **457**, 706–710 (2009).

[126] X. Li, W. Cai, J. An, S. Kim, J. Nah, D. Yang, R. Piner, A. Velamakanni, I. Jung, E. Tutuc, S. K. Banerjee, L. Colombo, and R. S. Ruoff, "Large-Area Synthesis of High-Quality and Uniform Graphene Films on Copper Foils," Science **324**, 1312–1314 (2009).

Chapter 3

Theory of linear diffraction by periodic structures: the rigorous coupled-wave analysis

3.1 Introduction

The rigorous coupled-wave analysis RCWA is a fundamental numerical method to calculate the interaction of electromagnetic waves with periodic structures and has been used throughout the last few decades with large success. Its mathematical formulation became more general as new applications of gratings demanded a more versatile numerical method for their modelling, and it became more accurate and efficient once the connection between the mathematical theory of Fourier series multiplication and the RCWA was better understood. The main milestones in the development of the RCWA to a general method for periodic optical structures shall be briefly outlined.

Original formulations of RCWA were used to study linear optical wave-propagation in 1D planar and binary gratings [1, 2], metallic [3] and multilayered structures [4–6]. Its initial formulation shows excellent convergence for transverse electric (TE) polarization, however suffers from poor convergence for transverse magnetic (TM) polarization. The reason for this substantially slower convergence was an incorrect factorisation of truncated products of periodic functions in the Fourier space. For 1D gratings, this problem was solved practically in Ref. [7] by employing the so called inverse factorisation rule, for which a rigorous justification was given in Ref. [8]. An attempt to solve the Fourier factorisation problem, which Ref. [9] adequately describes in the context of any differential grating method, for 2D gratings was proposed in Ref. [10]. This formulation guarantees the correct Fourier factorisation for coordinate system-aligned gratings; however, it still suffers from slow convergence for rounded grating structures. In Refs. [11, 12] this issue successfully addressed for straight walled gratings with arbitrary cross section, and progress on the Fourier factorisation for oblique structures was made in Refs. [13, 14].

In contrast to the well established theory of RCWA for linear optical gratings, there have been only few formulations of RCWA for nonlinear optics. References. [15] and [16] proposed extensions to RCWA for bulk second- and higher-harmonic generation, respectively, and anisotropic quadratically nonlinear media were treated in Ref. [17]. All these formulations employ the undepleted pump approximation: the calculations at the fundamental frequency (FF) are decoupled from the second- (or higher-) harmonic calculations. It was recently shown in Ref. [18] how to model the optical Kerr effect Eq. (2.37) in gratings using a self-consistent iteration of the linear RCWA. It has however yet to be shown, how to ensure the correct Fourier factorisation rules for the nonlinear part of nonlinear calculations in RCWA.

This chapter presents a derivation of the RCWA, which incorporates the relevant advances in its theory, which have just been outlined. In particular, it employs the normal vector field (NVF, [11]) approach for the solution of the Fourier factorisation problem and I will show how this NVF approach generalizes to oblique, 2D-periodic structures [19, 20]. This derivation also serves as a mathematical reference for the computer code I implemented for the photonics software company Photon Design Ltd. Details about the implementation and capabilities of this commercially available software package, `OmniSim RCWA` [21], will be given in Chapter 5. Additionally, this derivation will serve as the starting point for the theory of linear and nonlinear diffraction in periodically patterned 2D materials in Chapter 6.

The remainder of this chapter is organised as follows: Section 3.2 will introduce the the physical setting of the grating problem. In order to facilitate the derivation of the RCWA, Section 3.3 will give the necessary mathematical definitions and theorems regarding Fourier series. Section 3.4 will continue this derivation by reformulating Maxwell's equations into a coupled system of ordinary differential equations, and introduce the concept of RCWA modes. The staircase approximation used by RCWA and a numerically stable scattering matrix formalism for multilayered structures are introduced in Section 3.5, before some of the important algorithmic choices of the RCWA are illustrated by an example in Section 3.6. Final remarks and conclusions about the RCWA will be given in Section 3.7.

3.2 Physical setting: the grating problem

The geometric setting for the presentation of the RCWA method are introduced in this Section. Thus, consider the generic multilayered periodic structure presented in Fig. 3.1. The close-up view in Fig. 3.1 depicts the unit cell of a 2D-periodic structure with periods Λ_1 and Λ_2, with the corresponding grating vectors, $\vec{\mathbf{\Lambda}}_1 = \mathbf{e}_x\Lambda_1$ and $\vec{\mathbf{\Lambda}}_2 = \mathbf{e}_y\Lambda_2$, laying in the (x, y)-plane. It consists of several bulky, periodically structured regions which will be called bulk-layers, or simply layers. The shape of the layers and the isotropic materials they comprise are described by relative electric permittivity distribution, $\epsilon_r(\mathbf{r})$. For generality

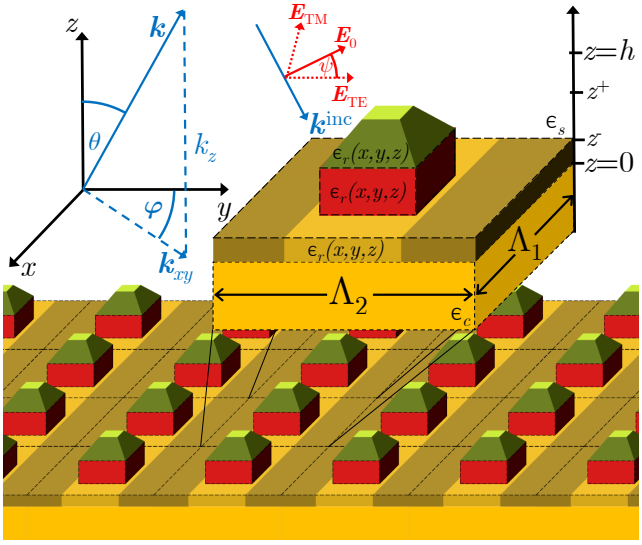

Figure 3.1: Schematic of a generic multilayer x, y-periodic structure under plane wave incidence, and a close-up of one unit cell. The specific shape of the grating is encoded in the permittivity distribution function $\epsilon_r(x, y, z)$, which allows the description of slanted (olive green), straight walled (red) or embedded or one dimensional parts (sand brown).

the depiction of the generic diffraction grating contains three different parts to illustrate specific aspects and components of diffraction gratings. The lowest part is a 1D-periodic binary grating structure consisting of alternating ribbons with different permittivity. Apart from that, a straight walled part with quadratic cross section, and a slanted grating, i.e. a grating part with oblique side walls, are depicted. The total periodic structure is enclosed in the semi-infinite homogeneous cover and substrate layers with relative permittivity, ϵ_c and ϵ_s, respectively. This is a specific illustration which should not distract from the fact that any diffraction grating structure can be described by the permittivity distribution $\epsilon_r(\mathbf{r})$.

This periodic structure is excited by an incident plane wave with the electric field given by:

$$\mathscr{E}^{\text{inc}}(\mathbf{r}, t) = \mathbf{E}^{\text{inc}}(\mathbf{r})e^{-i\omega t} = \mathbf{E}_0 e^{i(\mathbf{k}_t \cdot \mathbf{r}_t - k_z z - \omega t)} = \mathbf{E}_0 e^{i(\mathbf{k}^{\text{inc}} \cdot \mathbf{r} - \omega t)}, \tag{3.1}$$

where ω is the angular frequency, $\mathbf{r}_t$ is the in-plane component of the position vector, $\mathbf{k}_t = \mathbf{k}_t^{\text{inc}} = k_c(\sin\theta\cos\varphi, \sin\theta\sin\varphi)$ and $k_z^{\text{inc}} = -k_c\cos\theta$ are the in-plane and normal components of the wave vector of the incident wave in the cover, respectively, with $k_c = \sqrt{\epsilon_c}k_0 = \sqrt{\epsilon_c}\omega/c$ being the wavenumber in the cover region and c is the speed of light. The polarization state of the incoming plane-wave is described by the field components in the transverse electric and transverse magnetic configurations, namely $\mathbf{E}_{\text{TE}}$ and $\mathbf{E}_{\text{TM}}$, respectively, such that $\mathbf{E}_0 = \cos\psi\mathbf{E}_{\text{TE}} + \sin\psi\mathbf{E}_{\text{TM}}$ is determined by the polarization angle, ψ.

3.3 Mathematical background for Fourier series methods in photonics

Due to periodicity of the grating structure, the EM fields in and around the grating is pseudo-periodic, i.e. periodic with an additional transverse phase shift. It is hence natural to express the EM fields in terms of their Fourier series, which forms the basis of the numerical methods developed in this thesis, namely the RCWA and the GSM, see Chapter 7. Because a rigorous understanding of Fourier series is necessary to understand the key features of their numerical efficiency and convergence, the mathematical background of Fourier series will be given in this section.

3.3.1 Definition and basic properties of Fourier series

The definition, notation and a few well known theorems regarding trigonometric series are taken from Refs. [22, Chapters I and II] and [23, Chapter 16].

The Fourier series of a complex, 2π-periodic function f with $f(x) = f(x + 2\pi)$ is given by

$$S[f](x) = \sum_{n=-\infty}^{\infty} f_n \exp(inx) \tag{3.2}$$

where the coefficients f_n are defined by

$$f_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) \exp(-inx) dx. \tag{3.3}$$

The set of trigonometric functions $\{x \mapsto \exp(inx) : n \in \mathbb{Z}\}$ is orthogonal, in the sense that

$$\int_{-\pi}^{\pi} \exp(imx) \exp(-inx) dx = 2\pi\delta_{mn} := 2\pi \begin{cases} 0 & \text{if } m \neq n, \\ 1 & \text{if } m = n \end{cases} \tag{3.4}$$

for any $m, n \in \mathbb{Z}$. Therefore, these functions are also linearly independent, i.e. a vanishing Fourier series,

$$\sum_{n=-\infty}^{\infty} f_n \exp(inx) = 0 \text{ for all } x, \tag{3.5}$$

implies that all its coefficients vanish, $f_n = 0$ for all n.

Any integrable periodic function f, $\int_{-\pi}^{\pi} |f(x)|^2 dx < \infty$, has a Fourier series, i.e. the coefficients f_n in Eq. (3.3) exist. This, however, does not imply the way in which the Fourier

series $S[f]$ represents f. Results for different notions of convergence of the partial sums,

$$S_N[f](x) = \sum_{n=-N}^{N} f_n \exp(inx), \tag{3.6}$$

shall be briefly outlined, as they form the justification of the use of Fourier series as the basic building block of a numerical method. Thus, if $f(x)$ is piecewise continuous and piecewise continuously differentiable, $S_N(f)$ *converges pointwise* to f on $\mathbb{R} \setminus U_f$, where

$$U_f = \left\{ x \in \mathbb{R} : \lim_{\varepsilon \to 0} f(x+\varepsilon) - f(x-\varepsilon) \neq 0 \right\} \tag{3.7}$$

is the set of discontinuous points of f. This means

$$\lim_{N\to\infty} S_N[f](x) = f(x) \text{ for } x \in \mathbb{R} \setminus U_f. \tag{3.8}$$

For a point of discontinuity $x \in U_f$ however,

$$\lim_{N\to\infty} S_N[f](x) = \frac{1}{2} \lim_{\varepsilon\to 0} (f(x-\varepsilon) + f(x+\varepsilon)) \tag{3.9}$$

If $f(x)$ is continuous with piecewise continuous derivative, $S_N(f)$ *converges uniformly* to $\tilde{f}$ on $\mathbb{R}$, where $\tilde{f}(x) := \frac{1}{2} \lim_{\varepsilon\to 0}(f(x-\varepsilon) + f(x+\varepsilon))$. This means

$$\lim_{N\to\infty} \sup_{x\in\mathbb{R}} \left| S_N[f](x) - \tilde{f}(x) \right| = 0. \tag{3.10}$$

Uniform convergence implies pointwise convergence and is hence a stronger concept. Finally, L^2 *convergence*: If f is integrable, $S_N(f)$ converges to f in the L^2-norm:

$$\lim_{N\to\infty} \int_{-\pi}^{\pi} |S_N[f](x) - f(x)|^2 \, dx = 0. \tag{3.11}$$

These concepts of convergence will be important in the analysis of the RCWA and the GSM.

Other important basic properties of Fourier series are outlined here, and their proofs can be found in Ref. [22, Chapter II, Theorem 1.1 and Theorem 1.12]
Linearity of the S-operator: For periodic functions f and g, and $\alpha \in \mathbb{C}$:

$$S[f + \alpha g] = S[f] + \alpha S[g]. \tag{3.12}$$

Affine coordinate change: For constant $\lambda \in \mathbb{C}$ and $y \in \mathbb{R}$

$$f(\lambda x + y) \text{ is given by } \sum_{n=-\infty}^{\infty} f_n \exp(iny) \exp(in\lambda x). \tag{3.13}$$

Derivation rule: If f is differentiable

$$f'(x) \text{ is given by } i \sum_{n=-\infty}^{\infty} n f_n \exp(inx). \tag{3.14}$$

Parseval's formula: The norm of an integrable f is given by the l^2-norm of its coefficients

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x)|^2 dx = \sum_{n=-\infty}^{\infty} |f_n|^2. \tag{3.15}$$

One intrinsic feature of Fourier series is that all their terms $\exp(inx)$ are continuous, and in fact analytic. This leads to the well known Gibbs phenomenon, also known as ringing artefact in digital signal processing [24], which shall be illustrated now. Consider the rectangular function

$$f(x) = \begin{cases} 0 & \text{if } |x| > \pi/2, \\ 1 & \text{otherwise.} \end{cases} \tag{3.16}$$

on the interval $I = [-\pi, \pi)$ and its truncated Fourier series $S_N[f]$ with $N = 20$ as per Fig. 3.2(a). The partial sum $S_N[f]$ approximates f and in fact converges pointwise to f everywhere except for the points of discontinuity of $f(x)$. However, this convergence is not uniform but exhibits an over and undershooting behaviour around the discontinuities. This is highlighted in Fig. 3.2(b) for increasing N around $x_0 = -\pi/2$: whereas f is discontinuous and exhibits a jump of $\lim_{\varepsilon \to 0} f(x_0 + \varepsilon) - f(x_0 - \varepsilon) = 1$, all partial sums of $S[f]$ are continuous at x_0 . Additionally, the apparent over and undershoot of $S_N[f]$ near the discontinuity is given by

$$\lim_{N \to \infty} \max_{x \in I} S_N[f](x) - \min_{x \in I} S_N[f](x) = \frac{2}{\pi} \int_0^{\pi} \frac{\sin x}{x} dx = c_{\mathrm{G}} \approx 1.17898. \tag{3.17}$$

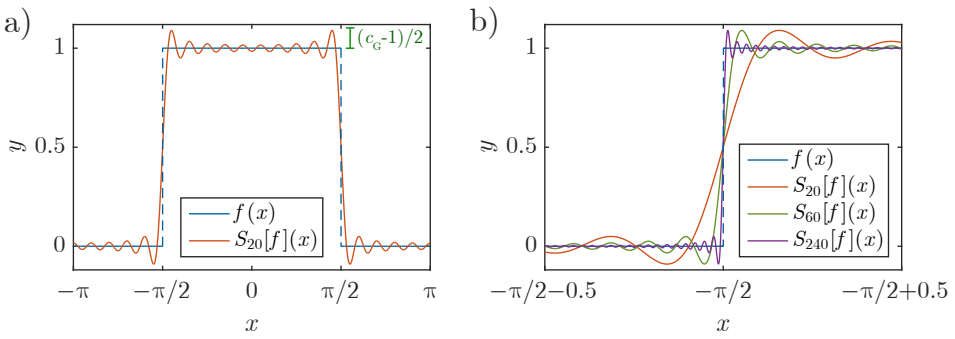

Figure 3.2: Illustration of the Gibbs phenomenon. a) Rectangular function and its Fourier series with $N = 20$ components. b) Close-up of region around point of discontinuity $x = -\frac{\pi}{2}$ highlighting the over- and undershoot due to the Gibbs phenomenon for increasing N.

The constant c_{G} is known as the Wilbraham-Gibbs constant, or simply Gibbs constant, named after Henry Wilbraham and Josiah Willard Gibbs. This problem, exemplified by the rectangular function, appears for any discontinuous function and carries over into numerical methods utilising Fourier series. However, with special precautions and care, it can be mostly circumvented in RCWA and the GSM, as will be shown later.

3.3.2 Fourier series in RCWA: Representation of the EM fields

The results from the previous section to 1D Fourier series can be generalised to Fourier series of two or more variables and can hence be used in the context of the RCWA.

The function of the permittivity distribution $\epsilon_r(x, y, z)$ of the structure in the (x-y)-plane is expressed by a 2D Fourier series

$$\epsilon_r(x, y, z) = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} \epsilon_{r,n}(z) \exp\left(2\pi i \left(\frac{n_1}{\Lambda_1}x + \frac{n_2}{\Lambda_2}y\right)\right) \tag{3.18}$$

with z-dependent coefficients $\epsilon_{r,n}(z)$ and $n = (n_1, n_2)$. In the following, the double sum $\sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty}$ will be denoted by $\sum_{n=-\infty}^{\infty}$. The Fourier series coefficients, $\epsilon_{r,n}(z)$, are given by

$$\epsilon_{r,n}(z) = \epsilon_{r,(n_1,n_2)}(z) = \frac{1}{\Lambda_1\Lambda_2} \int_0^{\Lambda_2} \int_0^{\Lambda_1} \epsilon_r(x, y, z) \exp\left(-2\pi i \left(\frac{n_1}{\Lambda_1}x + \frac{n_2}{\Lambda_2}y\right)\right) dydx. \tag{3.19}$$

Considering this structure under plane wave incidence from the top region as per Eq. (3.1), Bloch's theorem (Appendix A) guarantees the solution of ME is pseudo-periodic, i.e. periodic with an additional transverse phase shift. Therefore the EM field quantities $f = E_\alpha, H_\alpha, ..., \alpha = 1, 2, 3$, can be expressed as a Fourier series with phase shift,

$$f(x, y, z) = \mathscr{R}\left([f(z)]\right)(x, y) := \sum_{n=-\infty}^{\infty} f_n(z) \exp\left(i(k_{nx}x + k_{ny}y)\right) \tag{3.20}$$

where $k_{nx/y} = k_{0x/y} + n_{1/2}2\pi/\Lambda_{1/2}$ is the x/y component of the n-th diffraction order. The inherent phase shift is determined by the transverse component of the incident plane-wave k-vector $k_{0x/y} = k_{x/y}$. The sequence of Fourier series coefficients shall be denoted by $[f(z)]$, and the evaluation of a Fourier series by means of the infinite sum in Eq. (3.20) is denoted by the reconstruction operator $\mathscr{R}\left([f(z)]\right)$. This reconstruction operator incorporates the transverse phase shift $\exp(i(k_{0x}x + k_{0y}y)$.

The expansion of the EM fields in a periodic bulk layer of the grating, e.g. the middle part in Fig. 3.1, assumes a z-invariant permittivity function, $\epsilon_r(x, y, z) = \epsilon_r(x, y)$ for $z \in [z^+, z^-]$, where z^+ and z^- denote the bottom and the top of a periodic bulk layer. The way one can handle oblique layers, the top part in Fig. 3.1, will be described in Section 3.5. Due

to the z-independence of $\epsilon_r(x, y)$, also its Fourier series coefficients $\epsilon_{r,n}$ are z-independent. In contrast to that, the EM field quantities are given by z-dependent Fourier series and coefficients:

$$\mathbf{E}(\mathbf{r}) = \mathscr{R}\left([E_x(z)]\mathbf{e}_x + [E_y(z)]\mathbf{e}_y + [E_z(z)]\mathbf{e}_z\right) =: \mathscr{R}\left([\mathbf{E}(z)]\right) \tag{3.21a}$$

$$\mathbf{H}(\mathbf{r}) = \mathscr{R}\left([H_x(z)]\mathbf{e}_x + [H_y(z)]\mathbf{e}_y + [H_z(z)]\mathbf{e}_z\right) =: \mathscr{R}\left([\mathbf{H}(z)]\right). \tag{3.21b}$$

These expressions will be the starting point for the derivation of RCWA.

In actual numerical calculations, the infinite sums in Eqs. (3.18) and (3.21), need to be truncated. A rectangular truncation pattern of $n = (n_1, n_2) \in \{-N_1, \ldots, N_1\} \times \{-N_2, \ldots, N_2\}$ will be used throughout this derivation, yielding a total of $N_0 = (2N_1 + 1)(2N_2 + 1)$ Fourier series coefficients used in the representation of field quantities. The numbers N_1 and N_2 are often called the number of harmonics in the directions x and y, respectively, and N_0 can be called the total number of harmonics.

The truncation of products of periodic functions, in particular the constitutive relation

$$\mathbf{D}(x, y, z) = \epsilon_0 \epsilon_r(x, y)\mathbf{E}(x, y, z) \tag{3.22}$$

from Eqs. (2.2a) and (2.4a), will require special attention in order guarantee fast convergence of RCWA. Initial formulations of RCWA did not correctly address or ignored this issue, and hence exhibited slow convergence for TM polarization and 2D-periodic gratings. In fact, the Fourier factorisation problem not only arises in the context of RCWA, but also in numerical methods based on other series expansions of the solution, such as the Bessel-Fourier series or spherical harmonics for the scattering of light by cylindrical or arbitrary objects, respectively (see Refs. [25–27]). Due to its importance, the problem of Fourier factorisation will be introduced formally, before the correct treatment in the context of RCWA is presented.

3.3.3 The Fourier factorisation problem

This section will introduce the problem of Fourier factorisation in one dimension. It follows Refs. [8, 9] and borrows results from Refs. [22, 28]. Thereafter, solution strategies for 1D- and 2D-periodic grating structures are shown.

To start with, let $\mathbf{P}$ be the set of piecewise continuous, piecewise smooth, bounded, periodic functions in x with period 2π. For every $f \in \mathbf{P}$ and $g \in \mathbf{P}$, the function

$$h(x) = f(x)g(x) \tag{3.23}$$

is in $\mathbf{P}$ and the set of concurrent discontinuities of f and g is hence $U_{fg} = U_f \cap U_g$. Let the

jump of f, i.e. the amount of discontinuity of f at x, be defined as

$$\hat{f}_x := \lim_{\varepsilon \to 0} f(x+\varepsilon) - f(x-\varepsilon). \tag{3.24}$$

Note that $U_{fg} = U_h$ is not necessarily true, because of complementary discontinuities $x \in U_{fg}$ with $\hat{h}_x = 0$, i.e. $U_{fg} \supseteq U_h$.

The problem of Fourier factorisation can then be stated as:
Given the functions f and g, determine the coefficients h_n of their product h, without evaluating the product $f \cdot g$ in real space.
In case of untruncated Fourier series, this can be answered by Laurent's rule [22] [Chapter IV, Theorem 8.13], which yields an explicit representation of h_n:

$$h_n = \sum_{m=-\infty}^{\infty} f_{n-m} g_m \text{ for all } n. \tag{3.25}$$

If the number of terms in the Fourier series is truncated to $N > 0$, Laurent's rule results in coefficients

$$h_n^N = \sum_{m=-N}^{N} f_{n-m} g_m = \sum_{m=-N}^{N} [\![f]\!]_{nm} g_m \text{ for } n \in \{-N, \dots, N\} \tag{3.26}$$

which are in general different from h_n. Hereby, $[\![f]\!]_{nm} = f_{n-m}$ is the m,n entry of the $[\![f]\!] \in \mathbb{C}^{(2N+1)\times(2N+1)}$ with $4N+1$ independent entries. Matrices with the particular structure of the matrix $[\![f]\!]$ are called *Toeplitz* matrices and their properties and usage in the RCWA and GSM will be described in Appendix C. The hope is, that for increasing N, the partial sums obtained by the h_n and h_n^N converge uniformly to h and the Fourier factorisation problem is solved. This is not always the case and three situations can be distinguished, which are addressed by the theorems of Fourier factorisation. Their original formulation and discussion can be found in Lifeng Li's work [8], and their proofs in Ref. [29] [Appendix A].

Theorem I) If f and g have no concurrent jump discontinuities, $\lim_{N\to\infty} h_n^N = h_n$ for all n and hence $S_N[h^N]$ converges uniformly to h.

Theorem II) If f and g have concurrent jump discontinuities, application of Laurent's rule introduces an error:

$$S_n[h](x) = S_n[h^N](x) - \sum_{x_0 \in U_{fg}} \frac{\hat{f}_{x_0} \hat{g}_{x_0}}{2\pi^2} \Phi_N(x - x_0) + o(1) \tag{3.27}$$

where the term $o(1)$ converges uniformly to 0 as N increases and where the additional error

term is given by

$$\Phi_N(x) = \sum_{n=1}^{N} \frac{\cos nx}{n} \sum_{|m|>N} \frac{1}{m-n}. \tag{3.28}$$

This error function converges pointwise but nonuniformly to $\lim_{N\to\infty} \Phi_N(x) = \frac{\pi^2}{4}\delta(x)$.

Theorem III) Let S be a subinterval or a collection of subintervals of $[-\pi, \pi)$, and $\bar{S}$ its complement, where S or $\bar{S}$ can be empty, and assume $f(x) \neq 0$ and that all discontinuities of h are removable. If a) $\mathrm{Re}\,(1/f)$ does not change sign in $[-\pi, \pi)$, $\mathrm{Re}\,(1/f) \neq 0$ in S, and $\mathrm{Im}\,(1/f)$ does not change sign in $\bar{S}$ or b) $\mathrm{Im}\,(1/f)$ does not change sign in $[-\pi, \pi)$, $\mathrm{Im}\,(1/f) \neq 0$ in S, and $\mathrm{Re}\,(1/f)$ does not change sign in $\bar{S}$, then the partial sums $S_N[\tilde{h}_n^N]$ of the Fourier series defined by the coefficients

$$\tilde{h}_n^N = \sum_{m=-N}^{N} \left([\![1/f]\!]^{-1} \right)_{nm} g_m \text{ for } n \in \{-N, \ldots, N\} \tag{3.29}$$

converge uniformly to h. Here, $[\![1/f]\!]$ is the Toeplitz matrix of the Fourier series coefficients of $1/f(x)$, and hence Eq. (3.29) is called the *inverse rule*.

Theorems *I* and *III* directly give solutions to the Fourier factorisation problem, whereas Theorem *II* quantifies the error made when incorrect factorisation is performed. An example shall illustrate the important implications of correct and incorrect Fourier factorisation. Consider the functions

$$f(x) = \begin{cases} a, & |x| < \pi/2, \\ a/2, & |x| \geq \pi/2, \end{cases} \tag{3.30a}$$

$$g(x) = \begin{cases} b(1 - |x|/\pi), & |x| < \pi/2, \\ 2b(1 - |x|/\pi), & |x| \geq \pi/2, \end{cases} \tag{3.30b}$$

$$h(x) = f(x)g(x) = ab(1 - |x|/\pi) \tag{3.30c}$$

which are illustrated in Fig. 3.3(a) for the parameters $a = 3$ and $b = 2$. The functions f and g have complementary jump discontinuities at $|x| = \pi/2$ and their product h is continuous. Let the exact Fourier series coefficients of h be denoted by $[h]$, and the coefficients obtained by application of Laurent's rule Eq. (3.26) and the inverse rule Eq. (3.29) be given by $[h^L] = [\![f]\!][g]$ and $[h^i] = [\![1/f]\!]^{-1}[g]$, respectively. The partial sum $S_N[h^L]$ obtained by Laurent's rule exhibits strong oscillations near $|x| = \pi/2$. These oscillations are in fact given by $\Phi_N(|x| - \pi/2)$ due to Theorem *II*, and hence these oscillations do not vanish even for increasing N, as per Figs. 3.3(b) and 3.3(c) for $N = 10$ and $N = 200$, respectively. By contrast, the partial sums of $S_N[h^i]$ obtained by the inverse rule do not show spurious

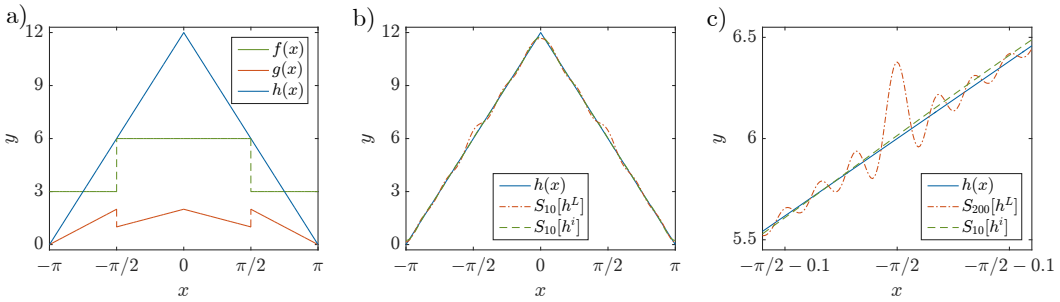

Figure 3.3: Illustration of correct and incorrect Fourier factorisation (adapted from Ref. [8]). a) Discontinuous functions f and g and their continuous product h. b) Fourier series reconstruction $S_N[\cdot]$ of coefficients $[h^L]$ and $[\tilde{h}^i]$, obtained by the Laurent's rule Eq. (3.26) and the inverse rule Eq. (3.29), respectively, for $N = 10$. c) A close-up around a point of discontinuity $x = -\pi/2$ reveals the oscillatory nature of $S_N[h^L]$ (for $N = 200$). In contrast to that, $S_N[h^i]$ converges uniformly to h.

oscillations and approximate h well even for small N.

This illustrates the main point of correct Fourier factorisation: Although the factors suffer from the Gibbs phenomenon, their product is free from numerical artefacts if the correct factorisation rule, i.e. the inverse rule, is used. The correct incorporation of these rules in the RCWA for 2D and 3D structures is called fast Fourier factorisation and will be demonstrated in the next sections.

3.3.4 Fast Fourier factorisation in 1D RCWA

Fast Fourier factorisation of 1D-periodic structures is straightforward with the rules just discussed, because the situations where Theorems *I* and *III* are applicable can be easily identified. To this end, consider a generic 1D-periodic grating described by $\epsilon_r(x, y) = \epsilon_r(x)$ and moreover, assume that $\epsilon_r(x) \neq 0$, and the additional preconditions of Theorem *III* are fulfilled, which is the case in gratings consisting of regular materials.

The permittivity distribution of a layer of any 1D grating structure can be defined on a set of N_S subintervals $S = \{S_i = [x_i, x_{i+1}) \subseteq \mathbb{R} : i = 1, ..., N_S - 1\}$, on which $\epsilon(x)$ is continuous. At each of the points x_i between the subintervals S_{i-1} and S_i, the permittivity $\epsilon_r(x)$ exhibits a jump $\hat{\epsilon}_r(x_i)$. This jump implies a discontinuity in the normal, i.e. x component of the electric field, E_x, yet the normal component of the displacement field, D_x, must be continuous due to boundary condition Eq. (2.6a). This commands the use of the inverse rule Eq. (3.29) for the Fourier factorisation of the x-components in Eq. (3.22):

$$[D_x(z)] = \epsilon_0 [\![1/\epsilon_r]\!]^{-1} [E_x(z)] \tag{3.31}$$

By contrast, the tangential components of $\mathbf{E}$, E_y and E_z, are continuous near x_i and the tangential displacement field is discontinuous. This allows direct application of Laurent's

rule Eq. (3.26) due to Theorem *I*:

$$[D_y(z)] = \epsilon_0 [\![\epsilon_r]\!] [E_y(z)] \tag{3.32a}$$

$$[D_z(z)] = \epsilon_0 [\![\epsilon_r]\!] [E_z(z)] \tag{3.32b}$$

3.3.5 Fast Fourier factorisation in 2D RCWA

Fast Fourier factorisation in 2D-periodic structures is more difficult, as the coordinate axis are typically not aligned with the interfaces between different materials, which rules out a straight forward application of Theorems *I* and *III* to the factorisation of the Cartesian components of Eq. (3.22). Three different approaches have been proposed and they will be introduced in the next sections.

3.3.5.1 Fast Fourier factorisation in 2D: Zig-zag approximation

The earliest rigorous approach to correct Fourier factorisation for 2D gratings consists of discretising an arbitrary surface contour Γ by a polygon Γ' consisting of straight lines which are parallel to the axes of the Cartesian coordinate system [10]. Figure 3.4(a) illustrates that normal and tangential directions of the discretised structure, Γ', can be easily distinguished, as these directions coincide with the basis directions of the coordinate system, x and y.

For 2D-periodic structures, the EM field quantities $\mathbf{E}$ and $\mathbf{D}$, and the permittivity ϵ_r are given as 2D Fourier series given in Eq. (3.21a) and Eq. (3.18), respectively. The idea is now to perform the 1D factorisation locally around Γ', where the continuity properties of $\mathbf{D}$ and $\mathbf{E}$ can be easily identified: At a point $\mathbf{r}'$ on a y-aligned line segment of Γ', D_x is continuous. Therefore, D_x is correctly factorised by applying the inverse rule with respect to x. The tangential E_y component, on the other hand, is continuous with respect to x, hence the y-component of Eq. (3.22) can be factorised by Laurent's rule with respect to x.

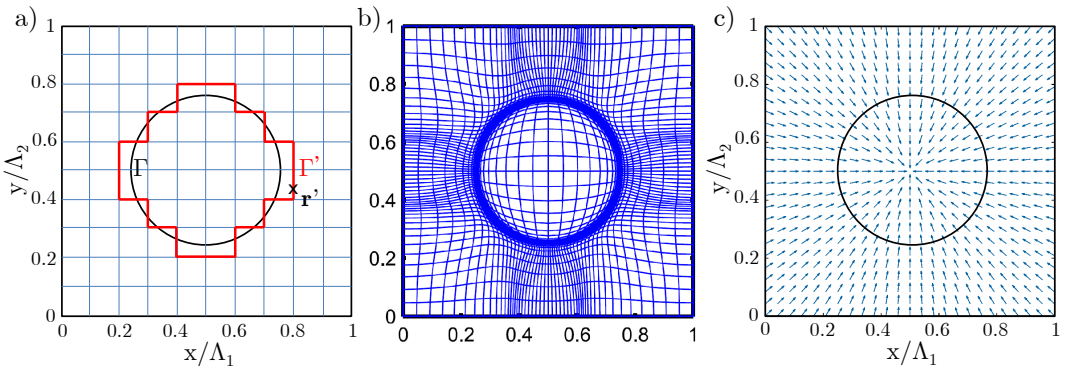

Figure 3.4: Factorisation strategies at the example of a cylindrical grating: a) *zig-zag* approach: The boundary Γ is approximated by a polygon Γ' consisting of coordinate system aligned straight lines. b) Coordinate transform [12]: The coordinate lines of the curvilinear coordinate system are aligned to the surface of the grating. c) normal vector field factorisation: The vector field is normal to the surface and continued into the unit cell.

Similar considerations for points on x-aligned segments of Γ' allow the formulation of the *zig-zag* approximation for Fourier factorisation in 2D.

$$[D_x(z)] = \lfloor\lceil\epsilon_r\rceil\rfloor[E_x(z)] \tag{3.33a}$$

$$[D_y(z)] = \lceil\lfloor\epsilon_r\rfloor\rceil[E_y(z)] \tag{3.33b}$$

$$[D_z(z)] = [\![\epsilon_r]\!][E_z(z)] \tag{3.33c}$$

The factorisation of the z-component of Eq. (3.22) is straight forward through Laurent's rule as z is tangential to Γ' and in fact to the original structure Γ itself. The matrix $[\![\epsilon_r]\!]$ in Eq. (3.33c) has therefore entries

$$[\![\epsilon_r]\!]_{n,m} = \epsilon_{r;n_1-m_1,n_2-m_2} \text{ for } n_i, m_i = -N_i, ..., N_i \text{ and } i = 1, 2. \tag{3.34}$$

$[\![\epsilon_r]\!] \in \mathbb{C}^{N_0 \times N_0}$ is hence a block Toeplitz matrix with $2N_1 + 1$ blocks, where each block itself is a Toeplitz matrix of size $2N_2 + 1$, see Appendix C. The notation of the matrices in Eqs. (3.33a) and (3.33b) requires additional explanations:

$$\lceil\epsilon_r(y)\rceil_{n_1,m_1} := \frac{1}{\Lambda_1}\int_0^{\Lambda_1} \epsilon_r(x,y)\exp\left(-i(n_2-m_2)\frac{2\pi}{\Lambda_1}x\right)dx, \tag{3.35a}$$

$$\lfloor\epsilon_r(x)\rfloor_{n_2,m_2} := \frac{1}{\Lambda_2}\int_0^{\Lambda_2} \epsilon_r(x,y)\exp\left(-i(n_1-m_1)\frac{2\pi}{\Lambda_2}y\right)dy \tag{3.35b}$$

are the entries of the Fourier series coefficient matrices for the x- and y-dependence of $\epsilon_r(x,y)$, respectively. Obviously, $\lceil\epsilon_r(y)\rceil_{n_1,m_1}$ and $\lfloor\epsilon_r(x)\rfloor_{n_1,m_1}$ are still functions of y and x, respectively. The final matrices of x- and y-independent coefficients used to factorise $[D_x]$ and $[D_y]$ are then given by

$$\begin{aligned}\lfloor\lceil\epsilon_r\rceil\rfloor_{n_1,n_2;m_1,m_2} &:= \left[\left\{\lceil 1/\epsilon_r(y)\rceil^{-1}\right\}_{n_1,m_1}\right]_{n_2,m_2} \qquad (3.36a)\\ &= \frac{1}{\Lambda_2}\int_0^{\Lambda_2}\left\{\lceil 1/\epsilon_r(\tilde{y})\rceil^{-1}\right\}_{n_1,m_1}\exp\left(-i(n_2-m_2)\frac{2\pi}{\Lambda_2}\tilde{y}\right)d\tilde{y},\\ \lceil\lfloor\epsilon_r\rfloor\rceil_{n_1,n_2;m_1,m_2} &:= \left[\left\{\lfloor 1/\epsilon_r(x)\rfloor^{-1}\right\}_{n_2,m_2}\right]_{n_1,m_1} \qquad (3.36b)\\ &= \frac{1}{\Lambda_1}\int_0^{\Lambda_1}\left\{\lfloor 1/\epsilon_r(\tilde{x})\rfloor^{-1}\right\}_{n_2,m_2}\exp\left(-i(n_1-m_1)\frac{2\pi}{\Lambda_1}\tilde{x}\right)d\tilde{x},\end{aligned}$$

respectively, where again $n_i, m_i = -N_i, ..., N_i$ for $i = 1, 2$. Note that the discretised interface polygon Γ' does not explicitly appear in Eqs. (3.35) and (3.36). It enters only at the stage of the numerical integration which is performed to calculate the coefficients of ϵ_r in the preprocessing stage of actual simulations. Therefore, the *zig-zag* approximation can be easily set up for any structure and is the correct factorisation if the structure is grid-aligned

($\Gamma = \Gamma'$). Moreover, it can be extended to the case of anisotropic materials and slanted coordinates [10, 30]. However, independently of how accurately the polygon Γ' approximates the contour Γ, the *zig-zag* approach yields slow convergence for diffraction gratings with rounded or arbitrarily shaped cross sections, which will be illustrated in Section 3.6.

3.3.5.2 Fast Fourier factorisation in 2D: Coordinate transformation

The coordinate transformation approach to fast Fourier factorisation takes advantage of the correctness of the *zig-zag* factorisation for grid-aligned structures. To this end, this approach seeks a coordinate transformation of the Cartesian coordinate system to a generally curvilinear system $O\bar{r}^1\bar{r}^2\bar{r}^3$ in which the boundaries of the structure are aligned to the new coordinate axes, as per Fig. 3.4(b). In the new coordinate system the material properties given by the effective permittivity $\bar{\epsilon}^{kl}(\bar{\mathbf{r}})$ and magnetic susceptibility $\bar{\mu}^{kl}(\bar{\mathbf{r}})$ in Eq. (2.21) are generally anisotropic, even when originally they are isotropic. Their possible discontinuities are however aligned to the curvilinear coordinate axes, whence the *zig-zag* factorisation for anisotropic materials, described in Ref. [30], is correct in $O\bar{r}^1\bar{r}^2\bar{r}^3$. Additionally, this approach allows for adaptive spatial resolution. This method has been shown to be accurate for both rectangular as well as rounded grating cross section and can be highly efficient. One main difficulty is the automated generation of an appropriate curvilinear coordinate system for arbitrary grating shapes. For the precise mathematical formulation refer to Refs. [12, 14, 31], as this approach is not further investigated here. Instead, an alternative treatment of the Fourier factorisation problem is considered throughout this thesis.

3.3.5.3 Fast Fourier factorisation in 2D: Normal vector field

The Fourier factorisation method mainly developed in this thesis is the normal vector field method [9, 11, 13, 20, 32]. The idea is to decompose the EM fields into their normal and tangential components at any point in the unit cell and then apply the correct factorisation rule to them according to Theorems *I* and *III*. This decomposition is performed by means of a vector field $\mathbf{N}(x, y)$, which is orthonormal to the grating surface everywhere on the interface contour, and is continuously extended into the entire of the unit cell, as per Fig. 3.4(c). This normal vector field can be found analytically for a few grating cross sections, and ways to automatically construct normal vector fields for arbitrary grating structures will be discussed in Section 5.3.2. Before the excellent convergence of the NVF approach for non-grid-aligned structures will be demonstrated in Section 3.6, the NVF factorisation is formulated exactly and the derivation of RCWA is continued in what follows.

Let $\mathbf{N}(x, y, z) = (N_x, N_y, N_z)^T(x, y)$ be the NVF in the grating layer under consideration. The NVF factorisation then combines three ideas: *i)* the EM fields $\mathbf{F} = \mathbf{D}$ and $\mathbf{E}$ are projected onto their normal and tangential component fields

$$\mathbf{F}_\perp(x, y, z) := \mathbf{N}(x, y)\mathbf{N}^T(x, y)\mathbf{F}(x, y, z) =: \mathbf{N}(x, y)F_\perp(x, y, z) \quad \text{and} \tag{3.37a}$$

$$\mathbf{F}_{\parallel}(x,y,z) := \left(\mathsf{I} - \mathbf{N}(x,y)\mathbf{N}^T(x,y)\right)\mathbf{F}(x,y,z), \tag{3.37b}$$

respectively. *ii)* By definition of the NVF, $\mathbf{D}_{\perp}(\mathbf{r})$ and $\mathbf{D}_{\parallel}(\mathbf{r})$ can be factorised by the inverse and Laurent's rule, respectively. Hereby, it is assumed that all discontinuities of $\mathbf{N}(x,y)$ are away from the surface contour Γ, otherwise, the prerequisites of Theorems *I* and *III* might not hold when applied to the fast Fourier factorisation of Eq. (3.37). This is not a strong limitation in practical scenarios. *iii)* The total displacement field is given by $\mathbf{D}(\mathbf{r}) = \mathbf{D}_{\perp}(\mathbf{r}) + \mathbf{D}_{\parallel}(\mathbf{r})$. Combining these ideas, one obtains the factorisation of the $\alpha = x, y, z$-component of $\mathbf{D}(\mathbf{r})$ in Eq. (3.22),

$$[D_\alpha(z)] = \epsilon_0 \sum_{\beta=1}^{3} \left(\delta_{\alpha,\beta}[\![\epsilon_r]\!] - \Delta\mathsf{N}_{\alpha\beta}\right)[E_\beta(z)], \tag{3.38}$$

where $\delta_{\alpha\beta}$ is the Kronecker delta. The matrix $\Delta\mathsf{N}_{\alpha\beta}$ is given by

$$\Delta\mathsf{N}_{\alpha\beta} = \frac{1}{2}\left(\Delta[\![N_\alpha N_\beta]\!] + [\![N_\alpha N_\beta]\!]\Delta\right) \tag{3.39}$$

with $\Delta = [\![\epsilon_r]\!] - [\![1/\epsilon_r]\!]^{-1}$ and $\alpha, \beta = x, y, z$.

Note that due to the commutation of Toeplitz operators [9], different, asymptotically equivalent definitions of $\Delta\mathsf{N}$ are possible. Most notably is a symmetric definition

$$\Delta\mathsf{N}_{\alpha\beta} = [\![N_\alpha]\!]\Delta[\![N_\beta]\!] \tag{3.40}$$

or one-sided definitions

$$\Delta\mathsf{N}_{\alpha\beta} = \Delta[\![N_\alpha N_\beta]\!], \tag{3.41a}$$

$$\Delta\mathsf{N}_{\alpha\beta} = [\![N_\alpha N_\beta]\!]\Delta. \tag{3.41b}$$

Because the Toeplitz matrices that arise due to the truncation of Fourier series do not commute, the different choices for the definition of $\Delta\mathsf{N}$ do influence the numerical results; see Section 3.6 for an illustration of this issue.

Before these results about the solution of the Fourier factorisation problem will be used to continue the derivation of the RCWA in the next section, it should be noted that it can be avoided by choosing a different set of basis function for the description of the EM fields. Most notably is the use of B-splines as basis functions, which accurately capture the behaviour of EM fields at the interfaces between media, see Refs. [33, 34].

3.4 The algebraic eigenvalue problem of RCWA

This section will introduce the general form of the eigenvalue problem of RCWA and the concept of a modal solution. The notation of Fourier series, their basic properties and the Theorems of Fourier factorisation introduced in the last section greatly simplifies this endeavour.

Let us start by inserting the periodic EM field expressions from Eq. (3.21) into the governing Maxwell's equations (2.9). Equation (3.38) is used to factorise $\mathbf{D}$ and since the complex exponentials of Fourier series are linearly independent, Ampere's law 2.9d can be rewritten in terms of the Fourier series coefficients $[E_\alpha(z)]$ and $[H_\alpha(z)]$

$$-\mathsf{K}_y[H_z(z)] - \frac{i}{k_0}\partial_z[H_y(z)] = \sum_{\beta=1}^{3}\left(\delta_{1,\beta}[\![\epsilon_r]\!] - \Delta\mathsf{N}_{1\beta}\right)[E_\beta(z)], \tag{3.42a}$$

$$\frac{i}{k_0}\partial_z[H_x(z)] + \mathsf{K}_x[H_z(z)] = \sum_{\beta=1}^{3}\left(\delta_{2,\beta}[\![\epsilon_r]\!] - \Delta\mathsf{N}_{2\beta}\right)[E_\beta(z)], \tag{3.42b}$$

$$-\mathsf{K}_x[H_y(z)] + \mathsf{K}_y[H_x(z)] = \sum_{\beta=1}^{3}\left(\delta_{3,\beta}[\![\epsilon_r]\!] - \Delta\mathsf{N}_{3\beta}\right)[E_\beta(z)] \tag{3.42c}$$

for $\alpha = x, y, z$. The matrix diagonal matrices K_x and K_y have entries k_{nx}/k_0 and k_{ny}/k_0, and they arise due to the transverse derivatives $\partial_x\mathbf{H}(x,y,z)$ and $\partial_y\mathbf{H}(x,y,z)$ in Eq. (2.9d), respectively. Similarly, one finds for Faraday's Law (2.9c)

$$\mathsf{K}_y[E_z(z)] - \frac{i}{k_0}\partial_z[E_y(z)] = [H_x(z)], \tag{3.43a}$$

$$-\frac{i}{k_0}\partial_z[E_x(z)] - \mathsf{K}_x[E_z(z)] = [H_y(z)], \tag{3.43b}$$

$$\mathsf{K}_x[E_y(z)] - \mathsf{K}_y[E_x(z)] = [H_z(z)]. \tag{3.43c}$$

The z-components E_z and H_z can be eliminated and a coupled system of ODEs for the z-dependent Fourier series coefficients is obtained:

$$\begin{aligned}-\frac{i}{k_0}\frac{\partial}{\partial z}[H_y(z)] =& -\Delta\mathsf{N}_{xz}\mathsf{B}\mathsf{K}_y[H_x(z)] + \Delta\mathsf{N}_{xz}\mathsf{B}\mathsf{K}_x[H_y(z)] \\ &+ \left(\mathsf{C}_x - \Delta\mathsf{N}_{xz}\mathsf{B}\Delta\mathsf{N}_{zx} - \mathsf{K}_y\mathsf{K}_y\right)[E_x(z)] + \left(\mathsf{K}_y\mathsf{K}_x - \Delta\mathsf{N}_{xz}\mathsf{B}\Delta\mathsf{N}_{zy} - \Delta\mathsf{N}_{xy}\right)[E_y(z)]\end{aligned} \tag{3.44a}$$

$$\begin{aligned}-\frac{i}{k_0}\frac{\partial}{\partial z}[H_x(z)] =& \Delta\mathsf{N}_{yz}\mathsf{B}\mathsf{K}_y[H_x(z)] - \Delta\mathsf{N}_{yz}\mathsf{B}\mathsf{K}_x[H_y(z)] \\ &+ \left(\Delta\mathsf{N}_{yx} + \Delta\mathsf{N}_{yz}\mathsf{B}\Delta\mathsf{N}_{zx} - \mathsf{K}_x\mathsf{K}_y\right)[E_x(z)] + \left(\mathsf{K}_x\mathsf{K}_x + \Delta\mathsf{N}_{yz}\mathsf{B}\Delta\mathsf{N}_{zy} - \mathsf{C}_y\right)[E_y(z)]\end{aligned} \tag{3.44b}$$

$$-\frac{i}{k_0}\frac{\partial}{\partial z}[E_y(z)] = \mathsf{K}_y\mathsf{B}\Delta\mathsf{N}_{zx}[E_x(z)] + \mathsf{K}_y\mathsf{B}\Delta\mathsf{N}_{zy}[E_y(z)] + \left(\mathsf{K}_y\mathsf{B}\mathsf{K}_y - \mathsf{I}\right)[H_x(z)] - \mathsf{K}_y\mathsf{B}\mathsf{K}_x[H_y(z)] \tag{3.44c}$$

$$-\frac{i}{k_0}\frac{\partial}{\partial z}[E_x(z)] = \mathsf{K}_x\mathsf{B}\Delta\mathsf{N}_{zx}[E_x(z)] + \mathsf{K}_x\mathsf{B}\Delta\mathsf{N}_{zy}[E_y(z)] + \mathsf{K}_x\mathsf{B}\mathsf{K}_y[H_x(z)] + (\mathsf{I} - \mathsf{K}_x\mathsf{B}\mathsf{K}_x)\,[H_y(z)] \tag{3.44d}$$

where $\mathsf{B} := (\llbracket\epsilon_r\rrbracket - \Delta\mathsf{N}_{zz})^{-1}$ and $\mathsf{C}_\alpha := \llbracket\epsilon_r\rrbracket - \Delta\mathsf{N}_{\alpha\alpha}$. Since the coefficient matrices of this system of ODEs are z-invariant, it can be solved by assuming complex exponential z-dependence of the solutions

$$[E_\alpha(z)] = [E_\alpha]\exp(ik_0\nu z) \tag{3.45a}$$

$$[H_\alpha(z)] = [H_\alpha]\exp(ik_0\nu z) \tag{3.45b}$$

with the complex propagation constant, ν. Inserting these relations into Eq. (3.44) one obtains the RCWA eigenvalue problem

$$\mathsf{M}\begin{pmatrix}[E_x]\\ [E_y]\\ [H_x]\\ [H_y]\end{pmatrix} = \nu\begin{pmatrix}[E_x]\\ [E_y]\\ [H_x]\\ [H_y]\end{pmatrix}, \tag{3.46}$$

where the system-matrix $\mathsf{M} \in \mathbb{C}^{4N_0\times 4N_0}$ form reads:

$$\mathsf{M} = \begin{pmatrix} \mathsf{K}_x\mathsf{B}\Delta\mathsf{N}_{zx} & \mathsf{K}_x\mathsf{B}\Delta\mathsf{N}_{zy} & \mathsf{K}_x\mathsf{B}\mathsf{K}_y & \mathsf{I} - \mathsf{K}_x\mathsf{B}\mathsf{K}_x \\ \mathsf{K}_y\mathsf{B}\Delta\mathsf{N}_{zx} & \mathsf{K}_y\mathsf{B}\Delta\mathsf{N}_{zy} & \mathsf{K}_y\mathsf{B}\mathsf{K}_y - \mathsf{I} & -\mathsf{K}_y\mathsf{K}_x \\ \Delta\mathsf{N}_{yx} + \Delta\mathsf{N}_{yz}\mathsf{B}\Delta\mathsf{N}_{zx} - \mathsf{K}_x\mathsf{K}_y & \mathsf{K}_x\mathsf{K}_x + \Delta\mathsf{N}_{yz}\mathsf{B}\Delta\mathsf{N}_{zy} - \mathsf{C}_y & \Delta\mathsf{N}_{yz}\mathsf{B}\mathsf{K}_y & -\Delta\mathsf{N}_{yz}\mathsf{B}\mathsf{K}_x \\ \mathsf{C}_x - \Delta\mathsf{N}_{xz}\mathsf{B}\Delta\mathsf{N}_{zx} - \mathsf{K}_y\mathsf{K}_y & \mathsf{K}_y\mathsf{K}_x - \Delta\mathsf{N}_{xz}\mathsf{B}\Delta\mathsf{N}_{zy} - \Delta\mathsf{N}_{xy} & -\Delta\mathsf{N}_{xz}\mathsf{B}\mathsf{K}_y & \Delta\mathsf{N}_{xz}\mathsf{B}\mathsf{K}_x \end{pmatrix}. \tag{3.47}$$

This is the most general form of the RCWA eigenvalue problem and no additional assumptions regarding the NVF were made. In particular, it allows the definition of an oblique NVF ($N_z \neq 0$), giving this formulation more generality than offered by previous derivations in the literature on the RCWA. Two practically important special cases shall be highlighted.

i) A 2D grating with straight walls under conical mount, i.e. $N_z = 0$, $\mathsf{B} = \llbracket\epsilon_r\rrbracket^{-1}$ allows Eq. (3.46) to be written as

$$\mathsf{M}_1\mathsf{M}_2\begin{pmatrix}[E_x]\\ [E_y]\end{pmatrix} = \nu^2\begin{pmatrix}[E_x]\\ [E_y]\end{pmatrix} \tag{3.48}$$

together with the relations for $[H_x]$ and $[H_y]$

$$\sqrt{\nu^2}\begin{pmatrix}[H_x]\\ [H_y]\end{pmatrix} = \mathsf{M}_1\begin{pmatrix}[E_x]\\ [E_y]\end{pmatrix} \tag{3.49}$$

where the $2N_0 \times 2N_0$ matrices M_i are given as block-matrices:

$$\mathsf{M}_1 = \begin{pmatrix} \mathsf{K}_x [\![\epsilon_r]\!]^{-1} \mathsf{K}_y & \mathsf{I} - \mathsf{K}_x [\![\epsilon_r]\!]^{-1} \mathsf{K}_x \\ \mathsf{K}_y [\![\epsilon_r]\!]^{-1} \mathsf{K}_y - \mathsf{I} & -\mathsf{K}_y [\![\epsilon_r]\!]^{-1} \mathsf{K}_x \end{pmatrix} \tag{3.50a}$$

$$\mathsf{M}_2 = \begin{pmatrix} \Delta \mathsf{N}_{yx} - \mathsf{K}_x \mathsf{K}_y & \mathsf{K}_x \mathsf{K}_x - \mathsf{C}_y \\ \mathsf{C}_x - \mathsf{K}_y \mathsf{K}_y & \mathsf{K}_y \mathsf{K}_x - \Delta \mathsf{N}_{xy} \end{pmatrix} \tag{3.50b}$$

This only requires the solution of two $2N_0 \times 2N_0$ eigenvalue problems and is the most commonly used formulation, even for gratings with oblique walls.

ii) A 1D grating with straight walls under non-conical mount, i.e. $N_x = 1$, $N_y = 0, N_z = 0$, $\mathsf{B} = [\![\epsilon_r]\!]^{-1}$, $\mathsf{C}_x = [\![1/\epsilon_r]\!]^{-1}$, $\mathsf{C}_y = [\![\epsilon_r]\!]$ and $\mathsf{K}_y = 0$, simplifies Eqs. (3.48) and (3.50) even further

$$\begin{pmatrix} [\![1/\epsilon_r]\!]^{-1} - \mathsf{K}_x [\![\epsilon_r]\!]^{-1} \mathsf{K}_y & [\![1/\epsilon_r]\!]^{-1} \\ 0 & [\![\epsilon_r]\!] - \mathsf{K}_x \mathsf{K}_y \end{pmatrix} \begin{pmatrix} [E_x] \\ [E_y] \end{pmatrix} = \nu^2 \begin{pmatrix} [E_x] \\ [E_y] \end{pmatrix}. \tag{3.51}$$

Therefore the TE (y) and TM (x) components of $\mathbf{E}$ decouple and only eigenvalue problems of size $N_o \times N_o$ need to be solved for the field component of interest. Note that Eq. (3.51) can be derived directly from the 1D factorisation rules introduced in Section 3.3.4.

The RCWA eigenproblem (3.46) has in general $4N_0$ solutions, which will be classified more specifically, after considering the case of a straight-walled layer, special case *i*. This version of the eigenvalue problem Eq. (3.48) has $2N_0$ solutions consisting of the eigenvalues ν_m^2 and eigenvectors $\left([E_x^m], [E_y^m]\right)$ for $m = 1, \ldots, 2N_0$. Defining the positive and negative roots of ν_m^2 as

$$\left(\nu_m^+\right)^2 := \nu_m^2 \text{ with } \mathrm{Re}\left(\nu_m^+\right) + \mathrm{Im}\left(\nu_m^+\right) > 0, \tag{3.52a}$$

$$\left(\nu_m^-\right)^2 := \nu_m^2 \text{ with } \mathrm{Re}\left(\nu_m^-\right) + \mathrm{Im}\left(\nu_m^-\right) \leq 0, \tag{3.52b}$$

respectively, one obtains a total of $2N_0$ upward and $2N_0$ downward propagating modes of the grating: The upward (downward) mode with number m is defined by its propagation constant ν_m^+ (ν_m^-) and its transverse mode profile given by the Fourier series coefficient vectors $[E_x^{(m,+)}], [E_y^{(m,+)}], [H_x^{(m,+)}], [H_y^{(m,+)}]$ $\left([E_x^{(m,-)}], [E_y^{(m,-)}], [H_x^{(m,-)}], [H_y^{(m,-)}]\right)$, where $[H_x^{(m,+)}]$ and $[H_y^{(m,+)}]$ $\left([H_x^{(m,-)}] \text{ and } [H_y^{(m,-)}]\right)$ are given by Eq. (3.49) with $\nu = \nu_m^+$ ($\nu = \nu_m^-$).

In the case of an oblique layer, it will be assumed that the $4N_0$ solutions of the general RCWA eigenproblem (3.46) can be also subdivided into $2N_0$ upward and $2N_0$ downward propagating modes. This assumption is not necessarily true in the general setting of an oblique grating layer. However, in most practical cases that were investigated, the assump-

tion of an identical number of upward as well as downward propagating modes still holds. If it is not fulfilled, however, the simplified matrices in Eqs. (3.50) should be used.

Since the bulk grating layer is considered to be made of linear materials, a linear superposition of modes is a solution to ME. Thus, the total electric field in the grating is given as

$$\begin{aligned} E_\alpha(\mathbf{r}) = \sum_{m=1}^{2N_0} & c_m^+ \mathscr{R}\left([E_\alpha^{(m,+)}]\right)(x,y) \exp\left(ik_0\nu_m^+(z-z^-)\right) \\ & + c_m^- \mathscr{R}\left([E_\alpha^{(m,-)}]\right)(x,y) \exp\left(ik_0\nu_m^-(z-z^+)\right) \end{aligned} \tag{3.53}$$

where the complex mode coefficient c_m^+ (c_m^-) determines the contribution of each upward (downward) propagating mode to the total grating field, and z^+ (z^-) denotes the bottom (top) z-coordinate of the considered grating layer. An analogous equation with the same coefficients $c_m^\pm$ determines $H_\alpha(\mathbf{r})$. Given the structural information of the modes, i.e. the mode profile and the propagation constant, the EM fields in the grating layer are hence fully determined by the $4N_0$ mode-coefficients $c_m^\pm$. Their actual values are obtained by means of the boundary conditions Eq. (2.6) for the electromagnetic fields at layer interfaces, which is treated in Section 3.5.

As it will be apparent later, it is useful to rewrite Eq. (3.53) in terms of z-dependent Fourier series coefficients, similarly to Eq. (3.21), but interchanging the summation order of modes and Fourier series components in Eq. (3.53):

$$E_\alpha(\mathbf{r}) = \mathscr{R}\left([E_\alpha^+(z)]\right)(x,y) + \mathscr{R}\left([E_\alpha^-(z)]\right)(x,y) \tag{3.54a}$$

$$H_\alpha(\mathbf{r}) = \mathscr{R}\left([H_\alpha^+(z)]\right)(x,y) + \mathscr{R}\left([H_\alpha^-(z)]\right)(x,y) \tag{3.54b}$$

where $[E^\pm(z)]$ and $[H^\pm(z)]$ are given by

$$\begin{aligned} [E_\alpha^\pm(z)] =& \sum_{m=1}^{2N_0} [E_\alpha^{(m,\pm)}(z)] = \sum_{m=1}^{2N_0} [E_\alpha^{(m,\pm)}] \exp\left(ik_0\nu_m^\mp(z-z^\mp)\right) c_m^\pm \\ =& \mathsf{E}_\alpha^\pm \mathsf{V}^\pm(z)\mathbf{c}^\pm, \end{aligned} \tag{3.55a}$$

$$[H_\alpha^\pm(z)] = \mathsf{H}_\alpha^\pm \mathsf{V}^\pm(z)\mathbf{c}^\pm. \tag{3.55b}$$

Here, the $N_0 \times 2N_0$ mode-shape matrix $\mathsf{E}_\alpha^\pm$ $\left(\mathsf{H}_\alpha^\pm\right)$ encodes the vector of Fourier series coefficients, $[E_\alpha^{(m,\pm)}]$ $\left([H_\alpha^{(m,\pm)}]\right)$, in its m-th column. The propagation matrix, $\mathsf{V}^\pm(z)$, is a diagonal matrix containing the z-dependence of each mode with $m \in \{1,\ldots,2N_0\}$ on its diagonal

$$\mathsf{V}_{mm}^\pm(z) = \exp\left(ik_0\nu_m^\pm(z-z^\mp)\right) \tag{3.56}$$

and $\mathbf{c}^{\pm}$ denotes the vector of upward ("+") and downward ("−") propagating mode coefficients.

The importance of Eq. (3.55) is that it translates the EM fields from their modal representation to their direct representation as Fourier series. The modal representation (r.h.s. of Eq. (3.55)) in terms of $\mathsf{E}_{\alpha}^{\pm}$, $\mathsf{V}^{\pm}$ and $\mathbf{c}^{\pm}$, is restricted to a computational layer, but very advantageous therein, as it separates mode-shape $\left([E_{\alpha}^{(m,\pm)}]\right)$, propagation or decay constant $(\nu_m^{\pm})$, and excitation strength $(c_m^{\pm})$. On the other hand, the direct representation as a Fourier series (l.h.s. Eq. (3.55)) with coefficients $[E_{\alpha}^{\pm}(z)]$ has the same complex exponentials, $\exp(ik_{nx}x + ik_{ny}y)$, as basis functions everywhere in the grating, and hence it facilitates the comparison of quantities in different computational layers.

3.5 $\mathcal{S}$-matrix formalism for multilayer structures

The derivation of the RCWA so far only considered one periodic layer, in which the RCWA modes Eq. (3.53) have been found. In an actual structure, this layer is not in isolation, but generally enclosed in additional layers, and the cover and substrate, as was illustrated in Fig. 3.1. Figure 3.5(a) shows the (y-z)-cross section through the centre of the unit cell of the grating. Whereas the EM fields in the straight parts of the grating are readily described by the RCWA formalism introduced in the last section, the slanted part deserves special attention because the permittivity distribution $\epsilon_r(x, y, z)$ is not z-independent therein. This problem is solved by using the staircase approximation [6] of slanted parts of a grating by approximating them with a stack of piecewise z-independent periodic layers, as is illustrated in Fig. 3.5(b). Each new computational slice that arises due to this approximation is assumed to be z-independent, and hence the RCWA modes therein can be found numerically by solution of Eq. (3.46).

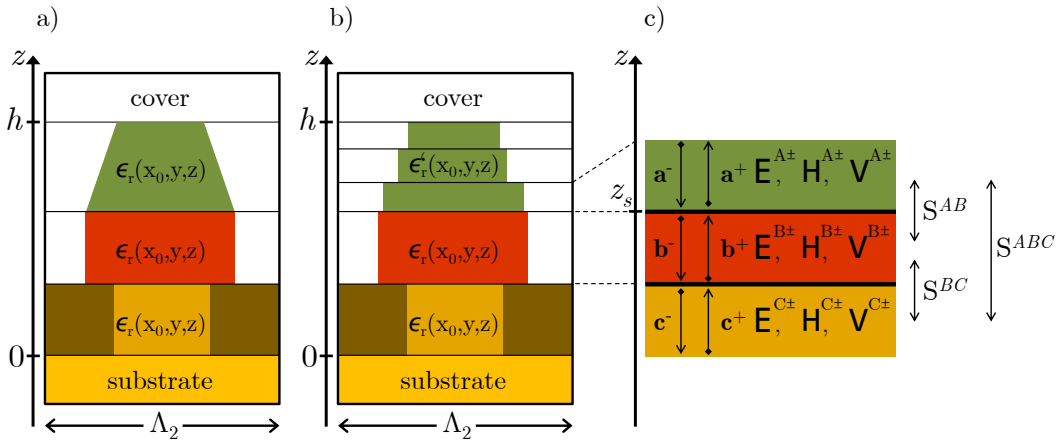

Figure 3.5: a) y-z-cross section of a generic grating diffraction grating with slanted walls. b) Discretisation of the slanted region by staircase the approximation. c) Setting for the $\mathcal{S}$-matrix formalism to connect modal coefficients $\mathbf{a}^{\pm}$, $\mathbf{b}^{\pm}$ and $\mathbf{c}^{\pm}$ in adjacent layers.

The EM fields in each computational layer are found by employing the EM boundary conditions at all interfaces between adjacent layers and different algorithmic approaches exist to match the modal fields in adjacent layers. The simplest is the transmission matrix approach which is in spite of its computational efficiency unsuitable for the RCWA; it is numerically unstable due to the potentially exponential growth of the complex modal propagation term $\exp(ik_0\nu z)$. Ways to overcome this instability of the transmission matrix approach in the RCWA are the enhanced transmittance matrix approach [4] and the scattering matrix ($\mathcal{S}$-matrix) formalism [5, 35].

This work makes use of the $\mathcal{S}$-matrix algorithm, to connect the mode coefficients of different computational layers and hence to determine the EM fields in the entire structure. The next section will introduce the $\mathcal{S}$-matrix formalism, which will be extended for the incorporation of linear and nonlinear 2D materials in Chapter 6.

3.5.1 $\mathcal{S}$-matrix of a two-layer interface system

The basis of the $\mathcal{S}$-matrix formalism is the modal representation of EM fields in any computational layer of the grating by Eq. (3.53). The last section has described this modal representation in periodic layers of the grating. In addition to that, the EM fields in the cover, substrate, and any other homogeneous layer inside the grating, can be described by Eq. (3.53). In this case, the modal representation of the EM fields is in fact given by a plane wave expansion, see Appendix B.

Consider now the two adjacent computational layers A and B in Fig. 3.5(c) and ignore the layer C for the moment. The modal EM fields in layers A and B are given in the compact matrix form Eq. (3.54), where the mode-shape matrices $\mathsf{E}_\alpha^{A,\pm}$ and $\mathsf{E}_\alpha^{B,\pm}$ are denoted with an additional superscript "A" for layer A and "B" for layer B. The notation for the propagation matrix $\mathsf{V}^{A/B,\pm}(z)$ is analogous and the modal coefficients in layers A and B are denoted by $\mathbf{a}^\pm$ and $\mathbf{b}^\pm$, respectively.

The aim of the $\mathcal{S}$-matrix formalism is to find the connection between $\mathbf{a}^\pm$ and $\mathbf{b}^\pm$ in the two layers such that the EM fields fulfil the EM boundary conditions Eq. (2.6) at the interface between the layers. In particular, the tangential components ($\mathbf{t} = (x, y)$) of $E_\mathbf{t}^{A/B}$ and $H_\mathbf{t}^{A/B}$ are continuous across the layer interface at $z = z_s$. Because $E_\mathbf{t}^{A/B}$ and $H_\mathbf{t}^{A/B}$ are given by their Fourier series coefficients, the tangential EM boundary conditions between the layers read:

$$[E_\mathbf{t}^A(z_s)] = [E_\mathbf{t}^B(z_s)], \tag{3.57a}$$

$$[H_\mathbf{t}^A(z_s)] = [H_\mathbf{t}^B(z_s)] \tag{3.57b}$$

Inserting the modal form of the EM fields Eq. (3.55) into Eq. (3.57) yields a matrix relation

between the coefficients

$$\mathsf{L}^{AB}\begin{pmatrix}\mathbf{a}^{+}\\ \mathbf{b}^{-}\end{pmatrix}=\mathsf{R}^{AB}\begin{pmatrix}\mathbf{b}^{+}\\ \mathbf{a}^{-}\end{pmatrix}, \tag{3.58}$$

where

$$\mathsf{L}^{AB}=\begin{bmatrix}\mathsf{E}_{\mathbf{t}}^{(A+)} & -\mathsf{E}_{\mathbf{t}}^{(B-)}\\ \mathsf{H}_{\mathbf{t}}^{(A+)} & -\mathsf{H}_{\mathbf{t}}^{(B-)}\end{bmatrix}\begin{bmatrix}\mathsf{V}^{(A+)}(z_s) & 0\\ 0 & \mathsf{V}^{(B-)}(z_s)\end{bmatrix}, \tag{3.59a}$$

$$\mathsf{R}^{AB}=\begin{bmatrix}\mathsf{E}_{\mathbf{t}}^{(B+)} & -\mathsf{E}_{\mathbf{t}}^{(A-)}\\ \mathsf{H}_{\mathbf{t}}^{(B+)} & -\mathsf{H}_{\mathbf{t}}^{(A-)}\end{bmatrix}\begin{bmatrix}\mathsf{V}^{(B+)}(z_s) & 0\\ 0 & \mathsf{V}^{(A-)}(z_s)\end{bmatrix} \tag{3.59b}$$

and $\mathsf{E}_{\mathbf{t}}^{L\pm}=\left[\mathsf{E}_x^{L\pm};\mathsf{E}_y^{L\pm}\right]$ denotes the $2N_0\times 2N_0$ matrix of all tangential Fourier series components of the modes in layer $L=A,B$. The matrix $\mathsf{H}_{\mathbf{t}}^{L\pm}=\left[\mathsf{H}_x^{L\pm};\mathsf{H}_y^{L\pm}\right]$ is defined in a similar way. A matrix inversion in Eq. (3.58) leads to the $\mathcal{S}$-matrix formalism of the two-layer interface system A-B:

$$\begin{pmatrix}\mathbf{a}^{+}\\ \mathbf{b}^{-}\end{pmatrix}=\left(\mathsf{L}^{AB}\right)^{-1}\mathsf{R}^{AB}\begin{pmatrix}\mathbf{b}^{+}\\ \mathbf{a}^{-}\end{pmatrix}=:\mathsf{S}^{AB}\begin{pmatrix}\mathbf{b}^{+}\\ \mathbf{a}^{-}\end{pmatrix}. \tag{3.60}$$

The matrix S^{AB} incorporates the upward and downward propagation of the incoming modes $\mathbf{b}^{+}$ and $\mathbf{a}^{-}$ in the layers B and A, respectively, and determines the coefficients $\mathbf{a}^{+}$ and $\mathbf{b}^{-}$ of the outgoing modes at the interface at z_s. This definition of S^{AB} is numerically stable, as the magnitude of all propagation terms $|\exp(ik_0\nu\cdot)|$ in $\mathsf{V}^{L\pm}(z_s)$ are ensured to be bounded by 1.

3.5.2 $\mathcal{S}$-matrix formalism for multilayer structures

It remains now to obtain the total $\mathcal{S}$-matrix of the diffraction grating, which connects the incoming coefficients in the cover with the coefficients of reflection and transmission in the substrate.

To formalize this, remember that electromagnetic fields in the homogeneous cover and substrate are expressed by the summation of the fields of the diffraction orders

$$\begin{aligned}\mathbf{E}^{C}(x,y,z)=&\mathbf{E}^{C+}(x,y,z)+\mathbf{E}^{C-}(x,y,z) &(3.61a)\\ =&\sum_{m=1}^{N_0}\left(\mathbf{E}_{m,e}^{C+}e_m^{C+}+\mathbf{E}_{m,h}^{C+}h_m^{C+}\right)\exp(ik_{mx}x+ik_{my}y+ik_{mz}^{C}(z-h))\\ &+\sum_{m=1}^{N_0}\left(\mathbf{E}_{m,e}^{C-}e_m^{C-}+\mathbf{E}_{m,h}^{C-}h_m^{C-}\right)\exp(ik_{mx}x+ik_{my}y-ik_{mz}^{C}(z+h)),\\ \mathbf{E}^{S}(x,y,z)=&\mathbf{E}^{S+}(x,y,z)+\mathbf{E}^{S-}(x,y,z) &(3.61b)\end{aligned}$$

$$= \sum_{m=1}^{N_0} \left(\mathbf{E}_{m,e}^{S+} e_m^{S+} + \mathbf{E}_{m,h}^{S+} h_m^{S+} \right) \exp(ik_{mx}x + ik_{my}y + ik_{mz}^S z)$$
$$+ \sum_{m=1}^{N_0} \left(\mathbf{E}_{m,e}^{S-} e_m^{S-} + \mathbf{E}_{m,h}^{S-} h_m^{S-} \right) \exp(ik_{mx}x + ik_{my}y - ik_{mz}^S z).$$

The incident field $\mathbf{E}^{C-}$ is hereby given by the TE and TM coefficients e_m^{C-} and h_m^{C-} and the TE and TM basis vectors $\mathbf{E}_{m,e}^{S+}$ and $\mathbf{E}_{m,h}^{S+}$ that are defined in Appendix B. The reflected field $\mathbf{E}^{C+}$ is defined by the coefficients e_m^{C+} and h_m^{C+} whereas the transmitted field $\mathbf{E}^{S-}$ is defined by coefficients e_m^{S-} and h_m^{S-}. Technically, this also allows for an incident field from the bottom of the structure, $\mathbf{E}^{S+}$, however it is considered to be zero in most practical problems. The tangential components of the wavevectors, $(k_{mx}, k_{my}) = (k_x + m_1 \frac{2\pi}{\Lambda_1}, k_y + m_2 \frac{2\pi}{\Lambda_2})$, are identical in cover and substrate, their z-components, however, it depends on the dispersion relation in the respective medium and are given by $k_{mz}^C = \sqrt{k_0^2 \epsilon_c - k_{mx}^2 - k_{my}^2}$ and $k_{mz}^S = \sqrt{k_0^2 \epsilon_s - k_{mx}^2 - k_{my}^2}$, respectively. The vector of all mode coefficients in the cover and substrate shall be denoted by $\mathbf{c}^{0,\pm} = \left(e_1^{C\pm}, h_1^{C\pm}, \ldots, e_{N_0}^{C\pm}, h_{N_0}^{C\pm} \right)$ and $\mathbf{c}^{N_L+1,\pm} = \left(e_1^{S\pm}, h_1^{S\pm}, \ldots, e_{N_0}^{S\pm}, h_{N_0}^{S\pm} \right)$, respectively, where N_L is the number of internal layers in the grating; in Fig. 3.5(b), $N_L = 5$. The magnetic fields in the cover and substrate are decomposed analogously. A comparison of Eq. (3.61) and Eq. (3.53) shows that the EM fields in cover and substrate can be cast in the same formalism of RCWA modes as the EM fields in any periodic layer. In contrast to the description of modes in a periodic layer, the field expansion in the homogeneous cover and substrate is given explicitly by Eq. (3.61) and does not require the solution of the RCWA eigenvalue problem.

The last section described the $\mathcal{S}$-matrix of two arbitrary adjacent layers. If one additionally considers a third layer, C, which is adjacent to B as per Fig. 3.5(c), the following $\mathcal{S}$-matrix relation holds for the mode coefficients therein:

$$\begin{pmatrix} \mathbf{b}^+ \\ \mathbf{c}^- \end{pmatrix} = \mathsf{S}^{BC} \begin{pmatrix} \mathbf{c}^+ \\ \mathbf{b}^- \end{pmatrix} \tag{3.62}$$

Straightforward matrix manipulations of Eqs. (3.60) and (3.62) result in expressions for the coefficients $\mathbf{b}^\pm$ in the intermediate layer B

$$\mathbf{b}^- = \left[\mathsf{I} - \mathsf{S}_{22}^{AB} \mathsf{S}_{11}^{BC} \right]^{-1} \left(\mathsf{S}_{21}^{AB} \mathbf{a}^- + \mathsf{S}_{22}^{AB} \mathsf{S}_{12}^{BC} \mathbf{c}^+ \right) \tag{3.63}$$
$$\mathbf{b}^+ = \left[\mathsf{I} - \mathsf{S}_{11}^{BC} \mathsf{S}_{22}^{AB} \right]^{-1} \left(\mathsf{S}_{11}^{BC} \mathsf{S}_{21}^{AB} \mathbf{a}^- + \mathsf{S}_{12}^{BC} \mathbf{c}^+ \right) \tag{3.64}$$

in terms of the coefficients of the incoming waves of neighbouring layers A and C. Hereby, S_{ij} denotes the $i, j = 1, 2$ submatrix of a matrix S of size $2N_0 \times 2N_0$. Similarly, one finds

the combined $\mathcal{S}$-matrix relation

$$\begin{pmatrix} \mathbf{a}^+ \\ \mathbf{c}^- \end{pmatrix} = \mathsf{S}^{ABC} \begin{pmatrix} \mathbf{c}^+ \\ \mathbf{a}^- \end{pmatrix} \tag{3.65}$$

where

$$\mathsf{S}^{ABC} = \mathsf{S}^{AB} \otimes \mathsf{S}^{BC} \tag{3.66}$$

denotes the Redheffer star-product [35] of two $\mathcal{S}$-matrices. It is given by

$$\mathsf{S}_{11}^{ABC} = \mathsf{S}_{11}^{AB} + \mathsf{S}_{12}^{AB} \left[\mathsf{I} - \mathsf{S}_{11}^{BC}\mathsf{S}_{22}^{AB}\right]^{-1} \mathsf{S}_{11}^{BC}\mathsf{S}_{21}^{AB}, \tag{3.67a}$$

$$\mathsf{S}_{12}^{ABC} = \mathsf{S}_{12}^{AB} \left[\mathsf{I} - \mathsf{S}_{11}^{C}\mathsf{S}_{22}^{AB}\right]^{-1} \mathsf{S}_{12}^{BC}, \tag{3.67b}$$

$$\mathsf{S}_{21}^{ABC} = \mathsf{S}_{21}^{BC} \left[\mathsf{I} - \mathsf{S}_{22}^{AB}\mathsf{S}_{11}^{BC}\right]^{-1} \mathsf{S}_{21}^{AB}, \tag{3.67c}$$

$$\mathsf{S}_{22}^{ABC} = \mathsf{S}_{22}^{BC} + \mathsf{S}_{21}^{BC} \left[\mathsf{I} - \mathsf{S}_{22}^{AB}\mathsf{S}_{11}^{BC}\right]^{-1} \mathsf{S}_{22}^{AB}\mathsf{S}_{12}^{BC}. \tag{3.67d}$$

The Redheffer star-product is non-commutative but associative and its neutral element is

$$\mathsf{I}^{\otimes} = \begin{pmatrix} 0 & \mathsf{I} \\ \mathsf{I} & 0 \end{pmatrix}. \tag{3.68}$$

The repeated application of Eq. (3.66) to the $\mathcal{S}$-matrices of all layers in the grating, including cover and substrate, yields the $\mathcal{S}$-matrix S of the total multilayered structure, which relates the cover coefficients, $\mathbf{c}^{0,\pm}$, and the substrate coefficients, $\mathbf{c}^{N_L+1,\pm}$, through

$$\begin{pmatrix} \mathbf{c}^{0,+} \\ \mathbf{c}^{N_L+1,-} \end{pmatrix} = \mathsf{S} \begin{pmatrix} \mathbf{c}^{N_L+1,+} \\ \mathbf{c}^{0,-} \end{pmatrix}. \tag{3.69}$$

Equation (3.69) describes the outgoing far-field, i.e reflection and transmission, in terms of the incident fields and Eq. (3.63) yields the mode coefficients in any computational layer. This allows evaluation of the EM fields everywhere in the grating by means of the Fourier series reconstruction Eq. (3.53) and with that the full solution of the grating problem is obtained.

From a practical point of view, the most important quantity derived from the solution coefficients is, apart from the EM fields in Eq. (3.61) itself, the normalised intensity of each diffraction order. It is customarily expressed as the ratio of the intensity of a diffraction order to the incident intensity, I_0, and called *diffraction efficiency*, DE. Thus, the intensities

of the single diffraction orders are given by

$$I_{me}^{\mathrm{ref/inc}} = \frac{1}{2}\sqrt{\frac{\epsilon_0}{\mu_0}}\mathrm{Re}\left(k_{mz}^{C}/k_0\right)|e_m^{C,\pm}|^2 \tag{3.70a}$$

$$I_{mh}^{\mathrm{ref/inc}} = \frac{1}{2}\sqrt{\frac{\epsilon_0}{\mu_0}}\mathrm{Re}\left(k_{mz}^{C}/k_0\right)|h_m^{C,\pm}|^2 \tag{3.70b}$$

$$I_{me}^{\mathrm{incS/trans}} = \frac{1}{2}\sqrt{\frac{\epsilon_0}{\mu_0}}\mathrm{Re}\left(k_{mz}^{S}/k_0\right)|e_m^{S,\pm}|^2 \tag{3.70c}$$

$$I_{mh}^{\mathrm{incS/trans}} = \frac{1}{2}\sqrt{\frac{\epsilon_0}{\mu_0}}\mathrm{Re}\left(k_{mz}^{S}/k_0\right)|h_m^{S,\pm}|^2 \tag{3.70d}$$

The subscript $m \in \{1, \ldots, N_0\}$ denotes the order of diffraction and the subscripts e an h distinguish between the TE and TM polarizations. The superscripts ref, inc and trans denote the reflected, incident and transmitted diffraction orders, respectively. The superscript incS denotes the diffraction orders incident from the substrate underneath the grating. Therefore, the total incident intensity is given by

$$I_0 = \sum_{m=1}^{N_0} \mathrm{I}_{me}^{\mathrm{inc}} + \mathrm{I}_{mh}^{\mathrm{inc}} + \mathrm{I}_{me}^{\mathrm{incS}} + \mathrm{I}_{mh}^{\mathrm{incS}} \tag{3.71}$$

and the reflected (ref) and transmitted (trans) TE (e) and TM (h) diffraction efficiency of order m is defined as

$$\mathrm{DE}_{me/h}^{\mathrm{ref/trans}} = \frac{I_{me/h}^{\mathrm{ref/trans}}}{I_0}. \tag{3.72}$$

The diffraction efficiencies are the defining quantities in classical applications of diffraction gratings, such as wavelength division multiplexing. From a numerical point of view, they are very suitable quantities for the investigation of the convergence of the RCWA, as they characterise the functionality of a diffraction gratings by a small number of numerical values.

If all materials in a grating are lossless, the optical power is conserved meaning that the power of the incident field is redistributed into the transmitted and reflected diffraction orders. With the definition of the diffraction efficiencies, the conservation of power for lossless grating structures can be formulated as

$$\sum_{m=1}^{N_0} \mathrm{DE}_m^{\mathrm{inc}} + \mathrm{DE}_m^{\mathrm{incS}} = \sum_{m=1}^{N_0} \mathrm{DE}_m^{\mathrm{ref}} + \mathrm{DE}_m^{\mathrm{trans}}, \tag{3.73}$$

where $\mathrm{DE}_m^{\cdot} = \mathrm{DE}_{me}^{\cdot} + \mathrm{DE}_{mh}^{\cdot}$ is the total, unpolarised diffraction efficiency of order m. In Ref. [36] it is demonstrated that the RCWA conserves power in the sense of Eq. (3.73) for 1D-periodic gratings for any number of harmonics N, and the application of the RCWA

in the years thereafter confirmed this for 2D-periodic gratings, too. This completes the derivation and discussion of the RCWA and a short instructive example should demonstrate the method now.

3.6 Application of the RCWA to a checkerboard grating

In this section, the factorisation rules for 2D-periodic gratings will be discussed using as an example a checkerboard grating structure, which has already been investigated in Refs. [10, 11, 37]. The unit cell of this grating is depicted in Fig. 3.6(a). It is placed on a substrate with relative permittivity $\epsilon_s = 1.5^2 = 2.25$ and consists of one periodic layer of height $h = \lambda$ and period $\Lambda_1 = \Lambda_2 = 1.25\lambda$. Each unit cell contains one cuboid with square cross section and transverse side length $a = \sqrt{2}\Lambda_1$. The material of the cuboid is given by $\epsilon_1 = \epsilon_s$, and the remaining parts of the grating and the cover are considered to be vacuum, $\epsilon_c = 1$. This structure is normally illuminated from the substrate ($\theta = 0$) with a plane wave with wavelength λ and the electric field is polarized along one of the sides of the square cross section.

Simulations are performed using both the *zig-zag* approximation 3.3.5.1 and the NVF approach 3.3.5.3 to the Fourier factorisation problem. The NVF $\mathbf{N}(x, y)$ used in the NVF factorisation is depicted in Fig. 3.6(b). As mentioned in Section 3.3.5.3 there are different possible definitions of the matrices $\Delta\mathsf{N}_{\alpha\beta}$, which discretise the relation $\mathbf{D} = \epsilon_0\epsilon_r\mathbf{E}$, and three choices are investigated here: the symmetric definition $\Delta\mathsf{N}_{\alpha\beta} = [\![N_\alpha]\!]\Delta[\![N_\beta]\!]$ (Eq. (3.40)), the one-sided definition $\Delta\mathsf{N}_{\alpha\beta} = \Delta[\![N_\alpha N_\beta]\!]$ (Eq. (3.41a)) and the symmetrised definition $\Delta\mathsf{N}_{\alpha\beta} = \frac{1}{2}\left(\Delta[\![N_\alpha N_\beta]\!] + [\![N_\alpha N_\beta]\!]\Delta\right)$ (Eq. (3.39)).

The diffraction efficiency of the central transmitted diffraction order, DE_0^T, is chosen as measure of convergence. Its value for calculations with different number of harmonics $N = N_1 = N_2 \in \{2, \ldots, 25\}$ is depicted in Fig. 3.6(c). The *zig-zag* approximation converges very slowly to the reference value ($\overline{\mathrm{DE}_0^T} = 0.1772$) and does not reach it even for $N = 25$, whereas all versions of the NVF factorisation rapidly converge to $\overline{\mathrm{DE}_0^T}$. The one-

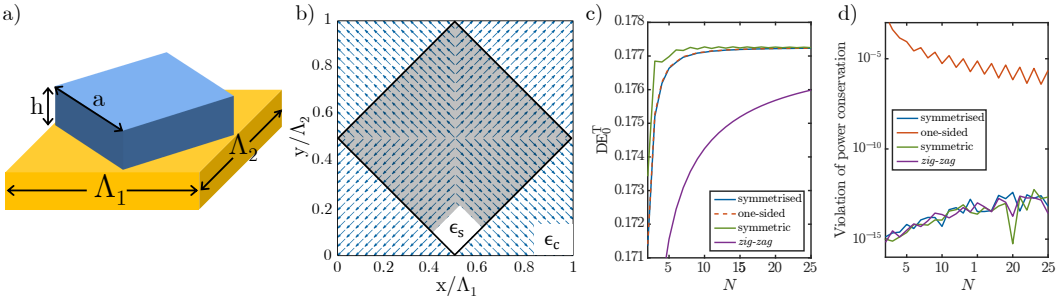

Figure 3.6: a) Schematic of unit cell of checkerboard grating. b) Normal vector field of the checkerboard grating. c) Convergence of th central transmitted diffraction efficiency DE_0^T with respect to N for different factorisation methods. d) Violation of the conservation of energy in term of an artificial absorption.

sided and symmetrised versions exhibit the same smooth convergence behaviour, whereas the symmetric form appears to reach $\overline{\mathrm{DE}}_0^T$ more rapidly. However, the convergence graph of the symmetric version oscillates around $\overline{\mathrm{DE}}_0^T$.

Because the grating structure only contains lossless media, all incident power should be conserved in the calculations. Figure 3.6(d) reveals that power is conserved up to the machine precision in the *zig-zag* approximation and the symmetric and symmetrised variations of the NVF approach. By contrast, the one-sided formulation only shows asymptotic conservation of power. This can be viewed both as a flaw of Eq. (3.41a), as well as an opportunity to check convergence of calculations without relying on a reference result. Two heuristic explanations of the smoother convergence of the one-sided or symmetrised versions can be given: *i)* the symmetric version $[\![N_\alpha]\!]\Delta[\![N_\alpha]\!]$ requires three successive multiplications of Fourier series matrices, whereas the one-sided or symmetrised versions versions only require two multiplications in $\Delta[\![N_\alpha N_\beta]\!]$. *ii)* $N_\alpha N_\alpha$ exhibits fewer discontinuities than N_α itself, and is hence more accurately described by its Fourier series.

Having investigated different versions of the NVF factorisation, an additional remark can be made about the similarities and differences between the factorisation approaches in two dimensions. The terms in Eq. (3.38) can formally be interpreted as the components of an anisotropic, effective medium tensor in Eq. (2.21), which indicates a strong connection between the NVF approach and the coordinate transform method from Section 3.3.5.2. Even more interestingly, Ref. [30, Section 5.2] finds the same ambiguity in the definition of effective $\bar{\epsilon}_r$ in the context of the coordinate system transformation approach, as it is found in the definitions of $\Delta\mathsf{N}$ here. However, despite the fact that the approaches seem similar, they are still different, as Li argues in Ref. [30, Section 5.2], by invoking the fundamental difference between inversion of ϵ in real space, $[\![1/\epsilon_r]\!]$, and inversion of the matrices of Fourier series coefficients, $[\![\epsilon_r]\!]^{-1}$. Additional to Li's arguments, I observed another strong difference between the two factorisation approaches: Symmetrising $\bar{\boldsymbol{\epsilon}}_r$ in the coordinate transformation factorisation spoils the nearly exact conservation of power, whereas only the symmetrised version of $\Delta\mathsf{N}$ in the NVF factorisation conserves the power. This opposite behaviour confirms the fundamental difference between the coordinate transform factorisation and the NVF factorisation.

3.7 Conclusions and remarks about the RCWA

This chapter has given the mathematical background and a exact formulation of the RCWA based on the NVF approach to fast Fourier factorisation and serves as a reference for the remaining chapters, in which modifications and improvements to the RCWA will be introduced. The NVF formulation has been put into the context of alternative formulations of RCWA and their similarity, and distinct differences, have been discussed. This chapter also formally introduced the general Fourier factorisation problem, which is relevant for

any numerical method based on Fourier series decomposition, illustrated it importance for accuracy and convergence of RCWA. While this chapter only contains the mathematical formulation of RCWA, the implementation details of the solution of the sub-problems, namely the Eigenvalue problem, the $\mathcal{S}$-matrix formalism, and the automated generation of NVF, will be deferred to Chapter 5.

This chapter introduces an important physical quantity, namely the notion of diffraction efficiency, which characterise the far-field behaviour of a diffraction grating and hence an elementary tool to describing the functionality of diffraction grating. Moreover, they are used to inspect convergence of RCWA calculations. One aspect that is not covered by the inspection of the diffraction efficiencies is the accuracy of the EM near-field inside a grating structure, which is of crucial importance for exact calculation of nonlinear optical effects in periodic structures. Therefore, the following chapter will introduce the accurate calculation of EM near-fields in diffraction gratings.

Bibliography

[1] M. G. Moharam and T. K. Gaylord, "Rigorous coupled-wave analysis of planar-grating diffraction," J. Opt. Soc. Am. **71**, 811–818 (1981).

[2] M. G. Moharam, T. K. Gaylord, 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).

[3] M. G. Moharam and T. K. Gaylord, "Rigorous coupled-wave analysis of metallic surface-relief gratings," J. Opt. Soc. Am. A. **3**, 1780–1787 (1986).

[4] M. G. Moharam, D. A. Pommet, E. B. Grann, and T. K. Gaylord, "Stable implementation of the rigorous coupled-wave analysis for surface-relief gratings: enhanced transmittance matrix approach," J. Opt. Soc. Am. A. **12**, 1077–1086 (1995).

[5] L. Li, "Formulation and comparison of two recursive matrix algorithms for modeling layered diffraction gratings," J. Opt. Soc. Am. A. **13**, 1024–1035 (1996).

[6] E. Popov, M. Neviere, B. Gralak, and G. Tayeb, "Staircase approximation validity for arbitrary-shaped gratings," J. Opt. Soc. Am. A. **19**, 33–42 (2002).

[7] P. Lalanne and G. M. Morris, "Highly improved convergence of the coupled-wave method for tm polarization," J. Opt. Soc. Am. A. **13**, 779–784 (1996).

[8] L. Li, "Use of fourier series in the analysis of discontinuous periodic structures," J. Opt. Soc. Am. A. **13**, 1870–1876 (1996).

[9] E. Popov and M. Neviere, "Maxwell equations in Fourier space: fast-converging formulation for diffraction by arbitrary shaped, periodic, anisotropic media," J. Opt. Soc. Am. A. **18**, 2886–2894 (2001).

[10] L. Li, "New formulation of the fourier modal method for crossed surface-relief gratings," J. Opt. Soc. Am. A. **14**, 2758–2767 (1997).

[11] 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**, 2880–2890 (2007).

[12] S. Essig and K. Busch, "Generation of adaptive coordinates and their use in the fourier modal method," Opt. Express **18**, 23258–23274 (2010).

[13] I. Gushchin and A. V. Tishchenko, "Fourier modal method for relief gratings with oblique boundary conditions," J. Opt. Soc. Am. A. **27**, 1575–1583 (2010).

[14] J. Küchenmeister, "Three-dimensional adaptive coordinate transformations for the fourier modal method," Opt. Express **22**, 1342–1349 (2014).

[15] W. Nakagawa, R.-C. Tyan, and Y. Fainman, "Analysis of enhanced second-harmonic generation in periodic nanostructures using modified rigorous coupled-wave analysis in the undepleted-pump approximation," J. Opt. Soc. Am. A. **19**, 1919–1928 (2002).

[16] T. Paul, C. Rockstuhl, and F. Lederer, "A numerical approach for analyzing higher harmonic generation in multilayer nanostructures," J. Opt. Soc. Am. B. **27**, 1118–1130 (2010).

[17] B. Bai and J. Turunen, "Fourier modal method for the analysis of second-harmonic generation in two-dimensionally periodic structures containing anisotropic materials," J. Opt. Soc. Am. B. **24**, 1105–1112 (2007).

[18] S. Bej, J. Tervo, Y. P. Svirko, and J. Turunen, "Modeling the optical Kerr effect in periodic structures by the linear Fourier modal method," J. Opt. Soc. Am. B. **31**, 2371–2378 (2014).

[19] M. Weismann and N. C. Panoiu, "Fourier modal method for efficient modeling of periodically structured metasurfaces with unit cells of arbitrary shape," in Advanced electromagnetic materials in microwaves and optics (metamaterials), 2014 8th international congress on (2014), pp. 289–291.

[20] M. Weismann, D. F. G. Gallagher, and N. C. Panoiu, "Accurate near-field evaluation in the rigorous coupled-wave analysis," J. Opt. **17**, 125612 (2015).

[21] Photon Design Ltd., *OmniSim/RCWA*, `http : / / www . photond . com / products/omnisim.htm`.

[22] A. Zygmund, *Trigonometric series*, Vol. 1 (Cambridge University Press, 2002).

[23] K. Königsberger, *Analysis 1*, 6th ed. (Springer Berlin Heidelberg, 2004).

[24] J. Chitode, *Digital signal processing* (Technical Publications, 2009).

[25] E. Popov, M. Nevière, and N. Bonod, "Factorization of products of discontinuous functions applied to fourier–bessel basis," J. Opt. Soc. Am. A. **21**, 46–52 (2004).

[26] B. Stout, M. Nevière, and E. Popov, "Light diffraction by a three-dimensional object: differential theory," J. Opt. Soc. Am. A. **22**, 2385–2404 (2005).

[27] N. Bonod, E. Popov, and M. Neviere, "Differential theory of diffraction by finite cylindrical objects," J. Opt. Soc. Am. A. **22**, 481–490 (2005).

[28] G. H. Hardy, *Divergent Series* (Oxford University Press, 1949).

[29] L. Li, "Mathematical reflections on the fourier modal method in grating theory," in *Mathematical modeling in optical science* (SIAM, Philadelphia, 2001) Chap. 4, pp. 111–139.

[30] L. Li, "Fourier modal method for crossed anisotropic gratings with arbitrary permittivity and permeability tensors," J. Opt. Soc. Am. A. **5**, 345 (2003).

[31] T. Weiss, G. Granet, N. A. Gippius, S. G. Tikhodeev, and H. Giessen, "Matched coordinates and adaptive spatial resolution in the Fourier modal method," Opt. Express **17**, 8051–8061 (2009).

[32] 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**, 1773–84 (2000).

[33] P. Bouchon, F. Pardo, R. Haïdar, and J.-L. Pelouard, "Fast modal method for subwavelength gratings based on b-spline formulation," J. Opt. Soc. Am. A. **27**, 696–702 (2010).

[34] M. Walz, T. Zebrowski, J. Küchenmeister, and K. Busch, "B-spline modal method: a polynomial approach compared to the fourier modal method," Opt. Express **21**, 14683–14697 (2013).

[35] R. C. Rumpf, "Improved formulation of scattering matrices for semi-analytical methods that is consistent with convention," Progress In Electromagnetics Research B **35**, 241–261 (2011).

[36] P. S. J. Russell, "Power conservation and field structures in uniform dielectric gratings," J. Opt. Soc. Am. A. **1**, 293–299 (1984).

[37] E. Noponen and J. Turunen, "Eigenmode method for electromagnetic synthesis of diffractive elements with three-dimensional profiles," J. Opt. Soc. Am. A. **11**, 2494–2502 (1994).

Chapter 4

Accurate near-field calculation in the rigorous coupled-wave analysis method

4.1 Introduction

The rigorous coupled-wave analysis method in the formulation of Chapter 3 has been successfully applied to model diffraction gratings, diffractive optical elements, surface coatings, spectroscopic applications, photonic crystals, and periodic metamaterials. Importantly, the functionality of these periodic photonic structures depends in most cases on the electromagnetic far-field, i.e. the field contained in the propagating diffraction orders. In this respect, it is known that the RCWA delivers fast converging and accurate far-field results, as reported in many works [1–11].

There is, however, a range of novel applications, which primarily rely on the optical near-field of a periodic structure, especially close to their surface, such as surface-enhanced Raman spectroscopy [12], surface second-harmonic generation [13–15], and near-field sensing [16–19]. These applications require a suitably designed distribution of the near-field, usually optimised for maximum field enhancement within specific spatial domains. Although there have been significant advances in experimental optical near-field measurement techniques [20], these techniques are still in their development stage and not readily available to accurately characterise complex photonic nanostructures. These applications and experimental shortcomings lead to a critical demand for numerical methods for periodic structures that can facilitate an accurate calculation of electromagnetic near-fields, and more importantly, the design of gratings with optimised near-field patterns.

With very few exceptions discussed in Refs. [7, 21, 22], a thorough investigation of numerically calculated near-fields in the RCWA has been largely overlooked during the development of the method, as it is often merely considered a post-processing step. Moreover, additional reasons for the scarcity of reports on the convergence and the accuracy of the numerically computed electromagnetic near-field in RCWA are the slow near-field convergence and spurious oscillations displayed by these fields, which is discussed in Ref. [22].

In this chapter I address the issue of inaccurate near-field calculations in the RCWA in several ways [23]. First, I use different generic cases of diffraction gratings to illustrate the slow convergence of RCWA for near-field calculations and reveal the reasons for this behaviour. Based on this analysis and the continuity properties of the electromagnetic fields, I will propose a new formulation of the field evaluation, which yields faster convergence of the electromagnetic near-fields and explicitly fulfils the continuity properties of the electromagnetic field at material interfaces. This improved formulation of the numerical evaluation of the near-fields is then benchmarked against the current formulation, with the aim of making RCWA an effective numerical method for modelling modern nanophotonic applications that rely on highly accurate near-field calculations.

The remainder of this chapter is organized as follows: Section 4.2 will introduce the reader to the problem of inaccurate and spurious near-fields in RCWA for 1D structures and will explain the overall strategy for overcoming these problems. Section 4.3 will extend the ideas gained from the analysis of 1D structures to arbitrary, straight or obliquely etched 2D gratings and will present the underlying mathematical formalism. Then in Section 4.4, computational results for 2D gratings will be presented and discussed. Section 4.5 will investigate near- and far-field convergence of the modified method for slanted gratings, before final conclusions about the capabilities of the improved RCWA for modelling electromagnetic near-fields will be drawn in Section 4.6.

4.2 Accurate near-field evaluation for 1D-periodic structures

Although the last major roadblock that was precluding RCWA from becoming a highly effective method for modelling 1D-periodic structures has been removed when the correct Fourier factorisation rules for transverse magnetic polarization were introduced [4, 5], a few topics have continued to attract attention, namely the convergence for slanted 1D-periodic gratings [11] and the numerical instabilities associated to highly-conductive gratings [24]. These improvements and refinements of the method, which are fully compatible with the formalism presented in Chapter 3, are concerned with the accuracy of the far-field calculation and they achieve excellent convergence results for the coefficients of and power carried by the diffraction orders in the cover and substrate regions of a grating, namely the far-field. This picture changes if the near-field is considered. Thus, in this section I investigate the numerically calculated EM near-field inside the periodic grating region via RCWA, and explain its slow convergence and the origin of the observed high spatial frequency oscillations. Thereafter I revisit a numerical method introduced in Ref. [21] for improved near-field accuracy and reduced numerical artefacts for straight 1D-periodic gratings. This numerical approach is thoroughly investigated here and will form the basis of the general formalism for accurate near-field evaluation in arbitrary, straight or slanted, 1D- and 2D-periodic gratings proposed in the subsequent sections.

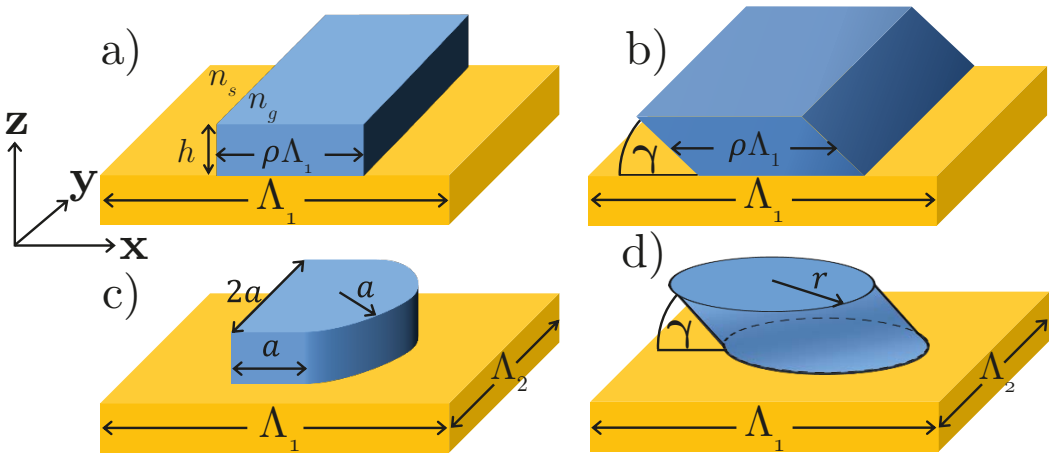

Figure 4.1: Grating structures mounted on a homogeneous substrate (refractive index, n_s) considered throughout this chapter: a) 1D binary grating, filling factor, ρ; b) 1D binary grating, slanted by $\gamma = \pi/4$ w.r.t. the x-axis. c) 2D grating with rectangle-semi-circular cross-section. d) 2D cylindrical grating, slanted by $\gamma = \pi/4$ w.r.t. the x-axis. The height of all gratings is denoted with h, their period is Λ_1 (and Λ_2 for 2D-periodic gratings), and their refractive index is denoted by n_g. The cover medium is vacuum with $n_c = 1$.

The ideas developed in this chapter are applicable to any periodic grating structure with sharp boundaries, but in order to conduct a comprehensive assessment of the accuracy of the near-field calculations achievable with the improved RCWA introduced in this chapter, a number of generic test structures (1D- and 2D-periodic, straight and slanted) have been chosen, as per Fig. 4.1. To further extend the spectrum of test configurations, three different materials are considered for the grating: silica (SiO_2) as a dielectric with low index of refraction ($n_g = n_{SiO_2} = 1.45$), silicon (Si) as a high-refractive index dielectric ($n_g = n_{Si} = 3.4$) and a metal, gold (Au), with index of refraction $n_g = n_{Au} = 0.97 + 1.87i$ [25] for a wavelength of $\lambda = 0.5\,\mu m$. In all examples the substrate is SiO_2 ($n_s = n_{SiO_2}$). Normal incidence and TM-polarization is considered in all of the examples, i.e. the incident electric field amplitude is oriented along the x-axis: $\mathbf{E}^{inc}(\mathbf{r}) = (1,0,0)^T E_0 \exp(-ik_c z)$ in Eq. (3.1).

The example considered in this section is depicted in Fig. 4.1(a). It consists of a binary grating with period, $\Lambda_1 = 1\,\mu m$, height, $h = 0.25\,\mu m$, and filling factor, $\rho = 0.5$. The grating is illuminated by a normally incident plane wave with wavelength, $\lambda = 0.51\,\mu m$.

As the first step of this investigation, RCWA is used to to calculate the far-field. In order to quantify the convergence of the RCWA, the relative error of the far-field is defined as:

$$e_F(N) = \frac{\sqrt{|T^{(N)} - T^{ref}|^2 + |R^{(N)} - R^{ref}|^2}}{\sqrt{|T^{ref}|^2 + |R^{ref}|^2}}, \tag{4.1}$$

where $T^{(N)}$ and $R^{(N)}$ denotes the relative transmitted and reflected power, respectively,

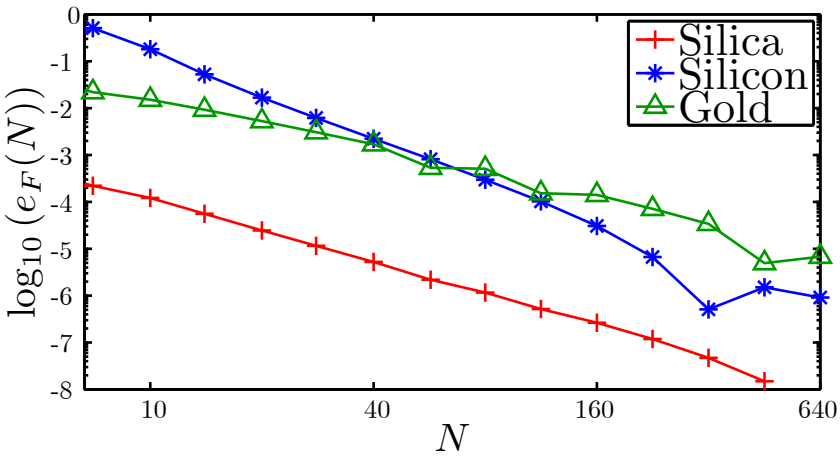

Figure 4.2: Far-field relative error vs. the number of harmonics, determined for three different materials.

corresponding to a discretisation with truncation number N in the Fourier series that describe the permittivity and the electromagnetic solution described in Section 3.3.2. More specifically, $(N_1, N_2) = (N, 0)$ and $(N_1, N_2) = (N, N)$ for 1D- and 2D-periodic gratings, respectively. Moreover, T^{ref} and R^{ref} are reference values that are considered to be the exact solution or a sufficiently good approximation of the exact values of the transmission and reflection, respectively. Due to the absence of an exact (analytical) solution of the diffraction grating problem, the reference values are chosen to be numerical values obtained by using a high value for N; in the case at hand, $T^{\text{ref}} = T^{(905)}$ and $R^{\text{ref}} = R^{(905)}$.

The relative far-field error, $e_F(N)$, for increasing number of harmonics, $N = 5, \ldots, 640$, is depicted in Fig. 4.2. As this figure illustrates, RCWA converges quickly for all three materials. Specifically, in order to achieve a self-error of $e_F(N) < 1\%$ (as a generic criterion adopted here for a satisfactorily accurate far-field calculation), for the three materials, silica (red crosses), silicon (blue stars), and gold (green triangles), a relatively small number of harmonics is necessary, namely $N > 5$, $N > 25$, and $N > 13$, respectively.

As mentioned in the introductory Section 4.1, the far-field of optical gratings and periodic structures has been the physical quantity of most interest from the experimental point of view, hence the characterization of a numerical method by means of the far-field convergence has usually been the adopted strategy. This approach, however, largely neglects the electromagnetic near-field predicted by a specific method. On the other hand, the near-field is of fundamental importance for modelling plasmonic effects or optical nonlinear phenomena in devices with size comparable to or smaller than the operating wavelength, effects whose description relies on accurate calculations of the electromagnetic near-field.

In order to characterise the numerically obtained near-fields, let us define the grating

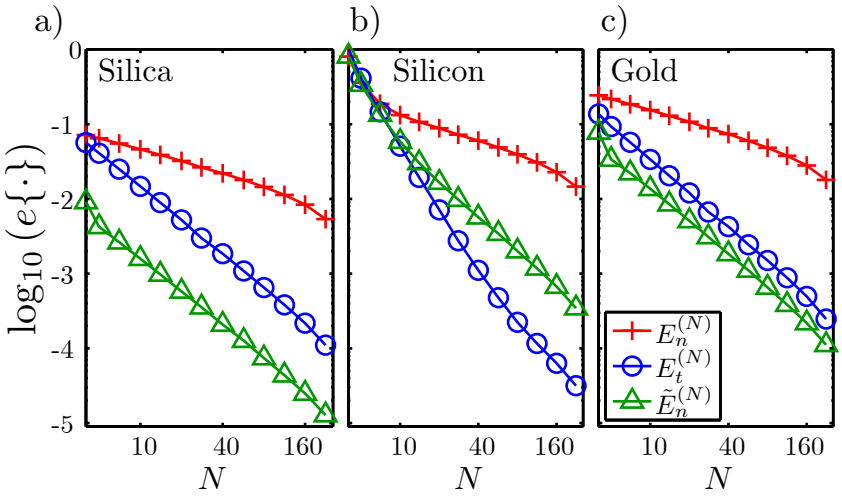

Figure 4.3: Self convergence of the tangential and normal electric field components E_z and E_x and $\tilde{E}_x$ inside the grating structure described by their self-errors $\Delta E_t(N)$ (green triangles), $\Delta E_n(N)$ (red crosses) and $\Delta \tilde{E}_n(N)$ (blue circles), respectively for the three benchmark structures made of Silica (a), Silicon (b) and Gold (c).

norm, $\|\cdot\|_G$, of a scalar or vectorial function, f, in the grating region as follows:

$$\|f\|_G = \left(\int_0^h \int_{-\Lambda_1/2}^{\Lambda_1/2} |f(x,z)|^2 \, dx \, dz \right)^{\frac{1}{2}}, \tag{4.2}$$

where the z-integration extends over the bulk of the periodic region. The grating norm is used to define the near-field error, $e\{E_\alpha^{(N)}\}$, of the scalar field components $E_\alpha^{(N)}$, $\alpha = x, y, z$, of the near-field, $\mathbf{E}^{(N)}$, which is numerically obtained using N harmonics:

$$e\{E_\alpha^{(N)}\} = \left\| E_\alpha^{(N)} - E_\alpha^{\text{ref}} \right\|_G \Big/ \left\| E_\alpha^{\text{ref}} \right\|_G. \tag{4.3}$$

Here, $E_\alpha^{(N)}$ denotes the Fourier series reconstruction of the electric field from Eq. (3.21a). In 1D-periodic gratings under normal incidence and for fixed $0 \leq z \leq h$, $E_\alpha^{(N)}(x,z)$ is explicitly given by the $2N+1$ Fourier coefficients, $E_{\alpha n}(z)$, which are calculated by RCWA:

$$E_\alpha^{(N)}(x,z) = \sum_{n=-N}^{N} E_{\alpha n}(z) \exp\left(in \frac{2\pi}{\Lambda_1} x \right). \tag{4.4}$$

Similarly to the far-field calculations, the reference near-field $\mathbf{E}^{\text{ref}} = \mathbf{E}^{(905)}$ is obtained by a high-resolution RCWA simulation with $N = 905$. The near-field self-convergence for the tangential component, $E_t = E_z$, and the normal component, $E_n = E_x$, of the electric field is depicted in Fig. 4.3. For all three materials, the self-convergence of E_t (blue circles) is fast and comparable to the far-field self-convergence (see Fig. 4.2). The

normal component, E_n (red crosses), however, exhibits much slower convergence and even at the highest numerical resolution of $N = 640$ a relative error of $e\{E_x^{(640)}\} > 0.9\%$ still remains, whereas the error of the tangential field, $e\{E_z^{(640)}\} < 0.08\%$, for all materials.

One intriguing question raised by the data plotted in Figs. 4.2 and 4.3 is why the normal component of the near-field converges much more slowly than the far-field power and the tangential component of the near-field. Or put it the other way around: how can the far-field converge quickly when the near-fields have not converged yet? There are three factors that explain this behaviour: *i)* The far-field consists of a superposition of a small number of propagating diffraction orders, i.e., plane waves. Hence, the far-field requires a small number of Fourier series components to be reconstructed. *ii)* The far field is defined in the homogeneous cover and substrate regions. It is therefore continuous in the transverse directions and does not suffer from the Gibbs phenomenon (Section 3.3.1). *iii)* RCWA does not depend on the representation of the discontinuous normal field E_n itself. Instead, the method relies on the correct Fourier factorisation of the continuous normal component of the displacement field, D_n, which can be accurately described by a Fourier series, i.e. without spurious oscillations, as was shown in Section 3.3.3. The third observation is at the core of the accurate near-field calculation that is introduced in this chapter. Specifically, instead of reconstructing a discontinuous physical quantity, i.e. $E_n(\mathbf{r})$, directly from its Fourier series coefficients, it is more effective to reconstruct a continuous quantity, the normal component of the displacement field, $D_n(\mathbf{r})$, and then divide it by a discontinuous quantity, the electric permittivity, $\epsilon_0\epsilon_r(\mathbf{r})$, which is known in the space-domain where one seeks to solve the diffraction problem.

Therefore one can define the modified normal component of the electric field for 1D-periodic gratings:

$$\tilde{E}_n^{(N)}(\mathbf{r}) = \epsilon_0^{-1} D_n^{(N)}(\mathbf{r})/\epsilon_r(\mathbf{r}), \tag{4.5}$$

where ϵ_0 is the vacuum permittivity and $\epsilon_r(\mathbf{r})$ the relative permittivity at position, $\mathbf{r}$. Note that $E_n^{(N)}(\mathbf{r})$ and $\tilde{E}_n^{(N)}(\mathbf{r})$ represent the same physical quantity, namely the normal component of the electric field. However, whereas the former is found by using RCWA to solve directly for the electric field, the latter one is determined by first calculating the displacement field and then the electric field via (4.5).

The error, $e\{\tilde{E}_n^{(N)}\}$, of the modified normal component, $\tilde{E}_n$, is shown in Fig. 4.3 and encoded in green triangles. It is found to converge as fast as the fast-converging tangential component, E_t, and as fast as the power in the far-field shown in Fig. 4.2. Even at the highest considered resolution, $N = 640$, the conventional formulation of the RCWA yields a self error of $e\{E_n^{(640)}\} > 9 \cdot 10^{-3}$ for all three materials. By contrast, the same self error $e\{\tilde{E}_n^{(N)}\} < 9 \cdot 10^{-3}$ of the modified normal field $\tilde{E}_n$ can be achieved by using as few as

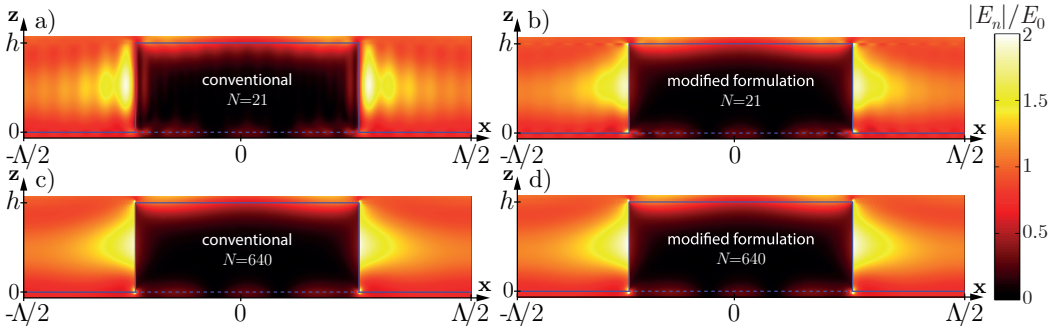

Figure 4.4: Near-field distribution of the normal component of the electric field, $|E_n| = |E_x|$, in and around the Au grating. a) Calculated with $N = 21$ harmonics using the conventional RCWA method (unphysical field oscillations can be observed). b) Calculated using the improved formulation, $\tilde{E}_n = D_n/\epsilon$, and $N = 21$ harmonics. c) Calculated using the conventional formulation and $N = 640$ harmonics. d) Calculated using the improved formulation and $N = 640$ harmonics.

$N = 70$ harmonics!

It should be noted, that this convergence study shows that the difference between conventional and modified formulation $\left\| E_n^{(N)} - \tilde{E}_n^{(N)} \right\|_G$ converges to 0. The error evolution, however, follows the trend of the slowly converging conventional formulation.

The spatial profile of the electric field in the grating region illustrates the full benefits of the modified field calculation. Figure 4.4(a) depicts the conventional normal field component $|E_n^{(21)}| = |E_x^{(21)}|$ in the gold grating for a moderately coarse discretisation of $N = 21$ harmonics. The field $|E_n^{(21)}|$ exhibits unphysical oscillations with a spatial frequency equal to the period of the smallest spatial frequency component in the Fourier series expansion of the solution. This is the well known Gibbs phenomenon, which occurs when describing a discontinuous function with a truncated Fourier series. On the other hand, the modified normal field, $\tilde{E}_n$, does not suffer from such spurious oscillations at the interface. In particular, even for a small number of harmonics, $N = 21$, $\tilde{E}_n^{(21)}$ is smooth, as per Fig. 4.4(b). At very large number of harmonics, $N = 640$, the modified normal field is free of any numerical artefacts, as can be seen in Fig. 4.4(d), and superficially agrees with the conventional field evaluation Fig. 4.4(c).

The improved formulation of RCWA exhibits another benefit, namely $\tilde{E}_n$ is by construction discontinuous and exactly fulfils the corresponding boundary condition,

$$\epsilon^{(\text{in})}\mathbf{E}^{(\text{in})}(\mathbf{r}_s) \cdot \mathbf{n} = \epsilon^{(\text{out})}\mathbf{E}^{(\text{out})}(\mathbf{r}_s) \cdot \mathbf{n}, \tag{4.6}$$

at surface points of the grating, $\mathbf{r}_s$, where $\mathbf{E}^{(\text{in})}$ and $\epsilon^{(\text{in})}$ ($\mathbf{E}^{(\text{out})}$ and $\epsilon^{(\text{out})}$) denote the electric field and permittivity inside (outside) the grating, respectively, and $\mathbf{n}$ is the unit vector normal to the surface. In the conventional formulation of the RCWA, the field $E_n^{(N)}$ does not satisfy Eq. (4.6), because $E_n^{(n)}$ is – as a Fourier series containing a *finite* number of

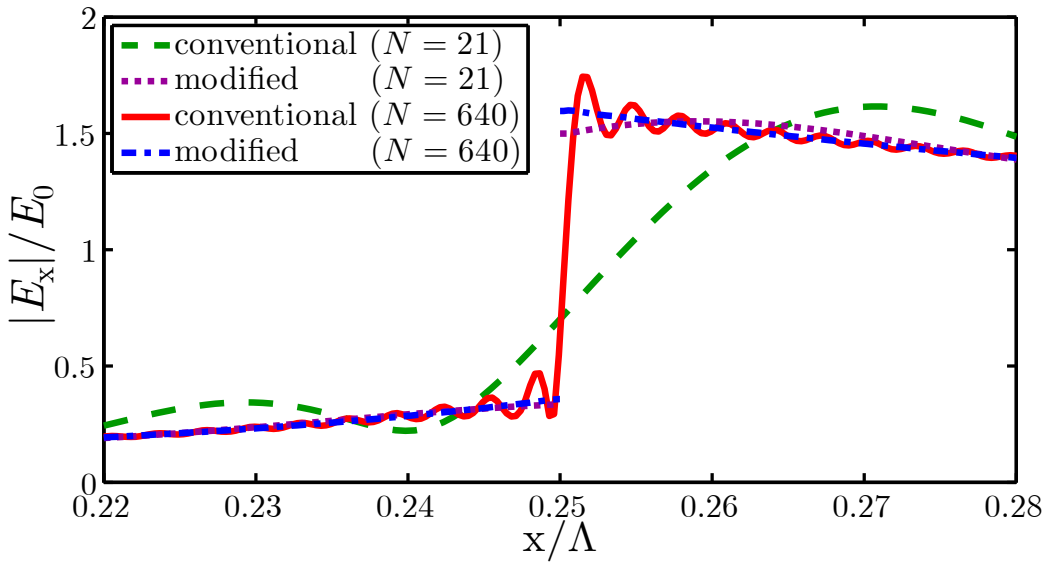

Figure 4.5: Close-up of the normal component of the electric field, $|E_n|$ and $|\tilde{E}_n|$, near the metal-air interface at $z = h/2$, computed using the conventional and modified RCWA methods, respectively.

terms – inherently continuous. The close-up of the interfacial field around $x = 0.25\Lambda$ in Fig. 4.5 emphasizes these ideas. Thus, the normal component of the field calculated using the conventional formulation of RCWA shows spurious oscillations for both small $N = 21$ (dashed green) and high $N = 640$ (solid red) number of harmonics, whereas the modified formulation is free of such unphysical oscillations and discontinuous for any number of harmonics, N (see the dotted purple and dashed-dotted blue lines corresponding to $N = 21$ and $N = 640$, respectively).

It should be clear now that the modified normal-field evaluation in 1D by means of the displacement field represents an improvement of the conventional evaluation of E_n in two ways: First, it exhibits optimal self-convergence in the sense that it is as accurate as the far-field and the continuous tangential field component. Second, it explicitly fulfils the boundary condition Eq. (4.6) at the material interfaces. Equally important, the improved near-field calculation is achieved with minimal additional computational cost and does not alter the mathematical framework of the core RCWA algorithm in 1D.

4.3 Formulation of accurate near-field evaluation for 2D-periodic structures

The ideas of the previous section are straightforward to implement because in 1D-periodic structures there is a trivial and unambiguous distinction between the tangential and normal components of **E**, which was already discussed in Section 3.3.4, where the fast Fourier factorisation for 1D gratings was introduced. This situation becomes more intricate in the

case of 2D-periodic diffraction gratings, where also the Fourier factorisation is a more complex problem. This section will generalise the concepts of the previous section and derive the formulation of accurate near-field calculations in 2D-periodic structures with straight or oblique walls.

The main aspect of the accurate near-field evaluation is that only the Fourier series of continuous quantities should be evaluated by means of their reconstruction operator $\mathscr{R}(\cdot)$ in Eq. (3.20), in order to avoid numerical artefacts. As we saw in Section 3.3.3, the continuity-properties of the EM fields are determined by the boundary conditions: E_t and D_n are continuous in the transverse plane, whereas D_t and E_n exhibit discontinuities at the interfaces between media, where n and t denote the normal and tangential components, respectively. In the NVF method for fast Fourier factorisation, this decomposition into normal and tangential components can be expressed by combining Eq. (3.38) and Eq. (3.37):

$$[(\mathbf{D}_n)_\alpha(z)] = \frac{1}{2}\epsilon_0 \sum_\beta \left([\![N_\alpha N_\beta]\!] [\![1/\epsilon_r]\!]^{-1} + [\![1/\epsilon_r]\!]^{-1} [\![N_\alpha N_\beta]\!] \right) [E_\beta(z)], \tag{4.7a}$$

$$[(\mathbf{E}_t)_\alpha(z)] = \sum_\beta [\![\delta_{\alpha\beta} - N_\alpha N_\beta]\!] [E_\beta(z)]. \tag{4.7b}$$

Hereby, $[(\mathbf{D}_n)_\alpha(z)]$ denotes the Fourier series coefficient vector of $(\mathbf{D}_n)_\alpha(x,y,z)$ at a level z in the grating. The field $(\mathbf{D}_n)_\alpha(x,y,z) = \mathbf{e}_\alpha \cdot \mathbf{D}_n(x,y,z)$ itself denotes the α-component of the normal displacement field $\mathbf{D}_n(x,y,z)$. Similarly, $[(\mathbf{E}_t)_\alpha(z)]$ denotes the Fourier series coefficient vector of the Cartesian α-component of the tangential electric field $\mathbf{E}_t(x,yz)$ at level z. This can be cast in a more concise matrix notation for the sums over the Cartesian coordinate β

$$[\mathbf{D}_n(z)] = \frac{1}{2}\epsilon_0 \left([\![\mathbf{N}\mathbf{N}^T]\!] [\![1/\epsilon_r]\!]^{-1} + [\![1/\epsilon_r]\!]^{-1} [\![\mathbf{N}\mathbf{N}^T]\!] \right) [\mathbf{E}(z)], \tag{4.8a}$$

$$[\mathbf{E}_t(z)] = [\![\mathsf{I} - \mathbf{N}\mathbf{N}^T]\!] [\mathbf{E}(z)]. \tag{4.8b}$$

Here the symmetrised definition of $\Delta\mathsf{N}_{\alpha\beta}$ (Eq. (3.39)) was used, but similar equations for Eq. (4.8a) can be formulated for the symmetric or one-sided definitions, Eqs. (3.40) and (3.41). By construction, $[\mathbf{D}_n]$ and $[\mathbf{E}_t]$ are Fourier series coefficient vectors of vector fields that are continuous at material interfaces. Therefore their reconstructions, $\mathscr{R}([\mathbf{D}_n])$ and $\mathscr{R}([\mathbf{E}_t])$, do not suffer from the Gibbs phenomenon at the interface. With this observation in mind, the electric field in the improved RCWA method for 2D-periodic structures at a point, $\mathbf{r} = (x,y,z)$, is given by:

$$\tilde{\mathbf{E}}^{(N)}(\mathbf{r}) = \epsilon_0^{-1} \mathscr{R}([\mathbf{D}_n(z)])(x,y)/\epsilon_r(\mathbf{r}) + \mathscr{R}([\mathbf{E}_t(z)])(x,y) \tag{4.9}$$

and is expected to yield faster near-field convergence and less oscillatory spatial field pro-

files as compared to the conventional field evaluation

$$\mathbf{E}^{(N)}(\mathbf{r}) = \mathscr{R}\left([\mathbf{E}(z)]\right)(x, y), \tag{4.10}$$

which does not take into account the continuity properties of the electric field. To investigate the validity of these predictions, the accurate field evaluation was implemented alongside the conventional formulation of RCWA, details of which will be discussed in Chapter 5.

It should be noted that the other electromagnetic fields can be easily calculated with this improved method, too. Specifically, the displacement field, $\mathbf{D}$, can be evaluated using the modified electric field $\tilde{E}$, namely $\mathbf{D}^{(N)}(\mathbf{r}) = \epsilon_0\epsilon(\mathbf{r})\tilde{\mathbf{E}}^{(N)}(\mathbf{r})$, and hence will have the same convergence properties as $\tilde{\mathbf{E}}$. The magnetic induction $\mathbf{B}$ and the magnetic field $\mathbf{H} = \mu\mathbf{B}$ do not require special attention, because they are continuous in non-magnetic materials and hence behave similarly to the continuous tangential component of the electric field.

4.4 Quantification of the accurate near-field evaluation in 2D-periodic structures

In this section, the improved formulation for accurate near-field calculations in 2D-periodic structures is assessed using two test structures with different configurations. To this end, the definition of the grating norm Eq. (4.2) of a scalar or vector function, f, is extended to 2D-periodic structures in the following straightforward way:

$$\|f\|_G = \left(\int_0^h \int_{-\Lambda_2/2}^{\Lambda_2/2} \int_{-\Lambda_1/2}^{\Lambda_1/2} |f(x, y, z)|^2 \, dx \, dy \, dz\right)^{\frac{1}{2}}, \tag{4.11}$$

where the integral is evaluated over the three-dimensional grating region.

4.4.1 Analysis of a 1D-periodic grating using 2D-RCWA

The first 2D-periodic diffraction grating under consideration is in fact a 1D binary grating, as shown in Fig. 4.1(a), rotated by $\pi/4$ in the x-y-plane. It should be obvious that it can be modelled as a double-periodic $2D$ grating with periods $\Lambda_1 = \Lambda_2 = \sqrt{2}\tilde{\Lambda} = \sqrt{2}\,\mu\mathrm{m}$, where $\tilde{\Lambda} = \Lambda_1$ is the period of the grating when it is viewed as a 1D-periodic structure. For clarity, the primary unit cell of the grating is depicted in the inset of Fig. 4.6(a). With this choice, the reference quantities $R^{\mathrm{ref}}, T^{\mathrm{ref}}$, and $\mathbf{E}^{\mathrm{ref}}$ can be accurately calculated using 1D simulations, an approach inspired by an example in Ref. [8].

The error in the calculation of the far-field, e_F from Eq. (4.1), when 2D simulations are employed is depicted in Fig. 4.6(a). The convergence of the calculations for the silicon and gold gratings closely follows the convergence trends observed when 1D simulations are performed and, as expected, it shows somewhat worse, yet still good, agreement for the silica structure. This difference is explained by the fact that the NVF introduced in the 2D-

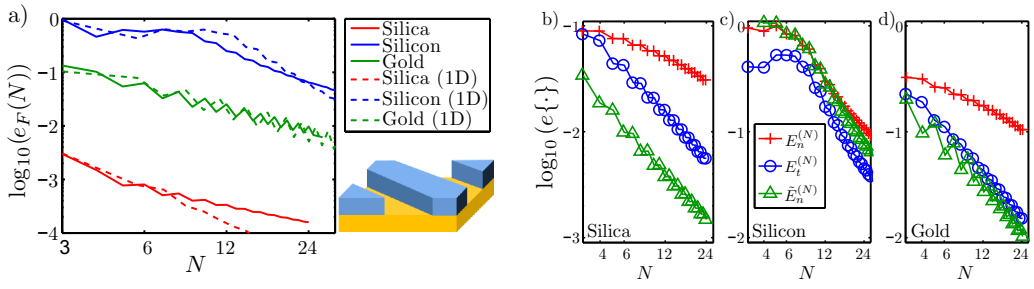

Figure 4.6: Computational results for a rotated binary 1D grating. a) Error of calculated far-field vs. N, determined for three gratings made of different materials. The decrease in the error follows that of the far-field (dashed lines) of the 1D simulations (see also section 4.2). b), c), d) Near-field error corresponding to the silica, silicon, and gold grating, respectively.

RCWA formulation is discontinuous away from material interfaces and hence it can degrade the convergence rate, especially for the low-index of refraction contrast case.

The conclusions of our analysis of the convergence of the near-field are summarised in Fig. 4.6(b-d). Thus, for both the silica and gold diffraction gratings, the normal component of the modified electric field, $\tilde{\mathbf{E}}$, exhibits faster convergence than this same field component calculated using the conventional form of RCWA. In both cases, $e\{\tilde{E}_n^{(20)}\}$ is one order of magnitude smaller than $e\{E_n^{(20)}\}$. For the silicon grating, only marginal differences between the two formulations can be observed. This is in agreement with the results obtained in the 1D case, as per Fig. 4.3(b). For small N, $\tilde{\mathbf{E}}^{(N)}$ and $\mathbf{E}^{(N)}$ are determined with a comparable degree of accuracy, but for the larger number of harmonics considered in Fig. 4.6(c), i.e. $N = 30$, a higher accuracy of our improved formulation of the RCWA can clearly be observed.

These conclusions are further validated by the profile of the electric field, as presented

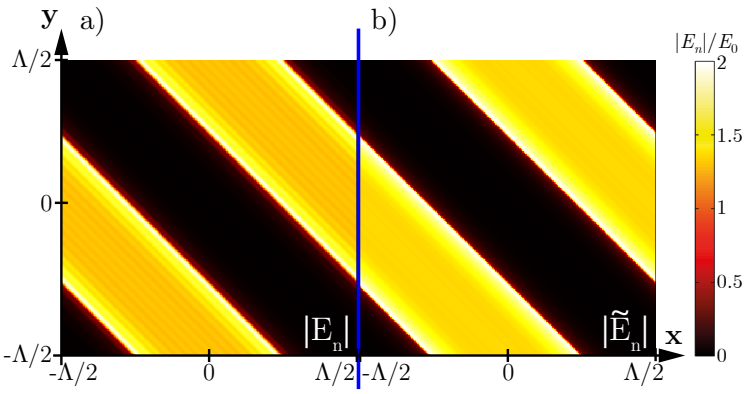

Figure 4.7: a) Normal component of the electric field, $|E_n|$, in the grating region at $z = h/2$, determined by using the conventional RCWA and $N = 27$. b) Normal electric field component, $|\tilde{E}_n|$, determined for the same grating parameters as in a) but using the improved algorithm. The blue vertical line was added for clarity and merely separates the two plots.

in Fig. 4.7. This figure shows the spatial distribution of the normal component of the electric fields, $\mathbf{E}^{(27)}$ and $\tilde{\mathbf{E}}^{(27)}$, calculated in the median plane of the grating. A simple examination of these field profiles confirms that the spurious oscillations of the field $\tilde{\mathbf{E}}^{(27)}$ near the surface have much smaller amplitude as compared to that of the variations of $\mathbf{E}^{(27)}$. Moreover, a closer inspection of the surface-fields shows that the boundary condition Eq. (4.6) is fulfilled by $\tilde{\mathbf{E}}^{(27)}$ only. This first test-case already reveals that the improved near-field evaluation is more accurate in the case of 2D-periodic structures, too.

4.4.2 Near-field calculations for an intrinsically 2D-periodic grating

In order to thoroughly test the near-field evaluation for 2D-periodic structures using the improved RCWA presented in this chapter, the following challenging test structure depicted in Fig. 4.1(c) is considered. The grating region consists of a coordinate system aligned parallelepiped with the length of the sides aligned to the x-, y-, and z-axis being $a = 0.5\Lambda$, $2a$, and h, respectively, placed adjacently to a semicircular cylinder with radius a and height h, with $\Lambda = \Lambda_1 = \Lambda_2 = 0.25\,\mu\text{m}$. The structure is illuminated normally by a x-polarised plane wave with wavelength $\lambda = 0.5\,\mu\text{m}$.

Since it is generally computationally time consuming to obtain high-accuracy solutions in the case of 2D-periodic structures and due to the fact that the higher the ratio Λ/λ and the refractive index, the more harmonics are necessary to achieve convergence [22], a relatively small period-to-wavelength ratio of $\Lambda/\lambda = 0.25\,\mu\text{m}/0.5\,\mu\text{m} = 0.5$ is chosen for this example.

As reference values in the definition of the far-field error, e_F from Eq. (4.1), $T^{\text{ref}} = T^{(31)}$ and $R^{\text{ref}} = R^{(31)}$, namely results obtained from simulations with $N = 31$ are chosen. The convergence of the far-field, $e_F(N)$, is shown in Fig. 4.8(a), where the far-field physical quantity considered are the transmission and reflection coefficients, T and R, respectively. As expected, the fastest convergence can be observed for the silica grating, because it has a low refractive index. The numerical error obtained for the gratings made of gold and silicon are one and two orders of magnitude larger than in the case of the silica grating,

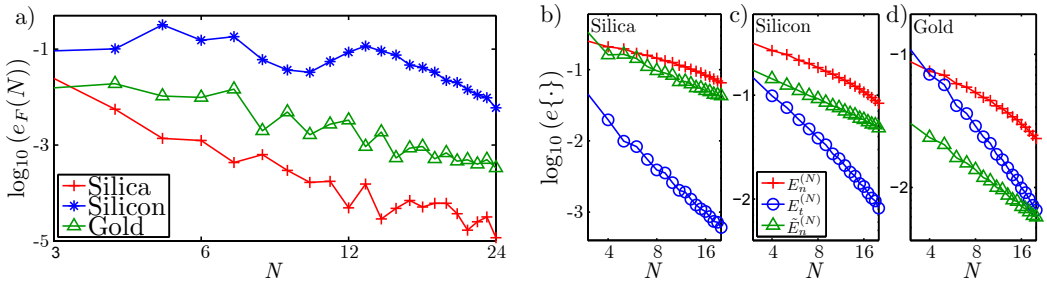

Figure 4.8: Computational results for the 2D grating shown in Fig. 4.1(c). a) Error of calculated far-field vs. N, determined for three gratings made of different materials. b), c), d) Near-field error corresponding to the silica, silicon, and gold grating, respectively.

respectively. This behaviour is similar to that seen in the 1D case for a small number of harmonics, $N < 30$ (cf. Fig. 4.2).

Figures 4.8(b)–4.8(d) contain the dependence of the near-field self-error determined for the three material configurations, silica, silicon, and gold, respectively. The three physical quantities plotted in each case are the z-component of $\mathbf{E}$ and the in-plane components, $\mathbf{E}_{xy} := (E_x, E_y, 0)^T$ and $\tilde{\mathbf{E}}_{xy} := (\tilde{E}_x, \tilde{E}_y, 0)^T$, obtained by using the conventional and improved versions of RCWA, respectively. Note that in either version, the in-plane component contains both the discontinuous normal component, $\mathbf{N} \cdot \mathbf{E}$ or $\mathbf{N} \cdot \tilde{\mathbf{E}}$, and the continuous tangential component, $(\mathsf{I} - \mathbf{N}\mathbf{N}^T)\mathbf{E}_{xy}$ or $(\mathsf{I} - \mathbf{N}\mathbf{N}^T)\tilde{\mathbf{E}}_{xy}$.

It can be seen that in all cases, the z-component of $\mathbf{E}$, which is continuous at vertical surfaces inside the grating region, converges much faster than the in-plane component, in both the conventional and modified formulations. In addition, the modified formulation leads to a somewhat smaller error than the conventional formulation. Specifically, it was found that $e\{\tilde{\mathbf{E}}_{xy}^{(N)}\} \approx 0.5e\{\mathbf{E}_{xy}^{(N)}\}$ for the silica grating and $e\{\tilde{\mathbf{E}}_{xy}^{(N)}\} \approx 0.9e\{\mathbf{E}_{xy}^{(N)}\}$ for the silicon and gold gratings. However, the corresponding convergence speed, i.e. the slope of $e\{\mathbf{E}_{xy}^{(N)}\}$ and $e\{\tilde{\mathbf{E}}_{xy}^{(N)}\}$, is the same. Moreover, in the modified formulation of the RCWA the convergence speed of the tangential component of the near-field is larger than that of the in-plane component. Three factors contribute to this behaviour: *i)* The decomposition of the near-field in a normal and tangential component in 2D-periodic structures relies on the specific definition of the normal vector field, $\mathbf{N}$ and it is not directly performed in Cartesian coordinates as in the 1D case. Hence, the inexact field decomposition by $\mathbf{N}$ introduces an additional error. *ii)* The field of normal vectors characterizing the structure is only uniquely defined at the interfaces defining the grating, except at the corners, and, more importantly, away from the grating surface. This ambiguity can lead to choices of NVFs which are not optimal for the convergence and accurate calculation of the near-field. *iii)* The normal vector field itself has discontinuities, which can cause additional oscillations in the spatial profile of the electromagnetic field.

It is also worthwhile to investigate the spatial profile of the near-field. The dominant x-component of the electric field in a horizontal cross-section through the grating region at $z = h/2$ is depicted in Fig. 4.9. As in the 1D case, these maps show that the field $\tilde{E}_x$ exhibits spatial oscillations with smaller amplitude as compared to the variations of the field E_x, especially at y-aligned interfaces (outlined with blue dashed lines in Fig. 4.9).

4.5 Out-of-plane normal vector fields for oblique diffraction gratings

Oblique diffraction gratings, as those shown in Figs. 4.1(b) and 4.1(d), are modelled in the RCWA within the staircase approximation along a direction perpendicular onto the grating

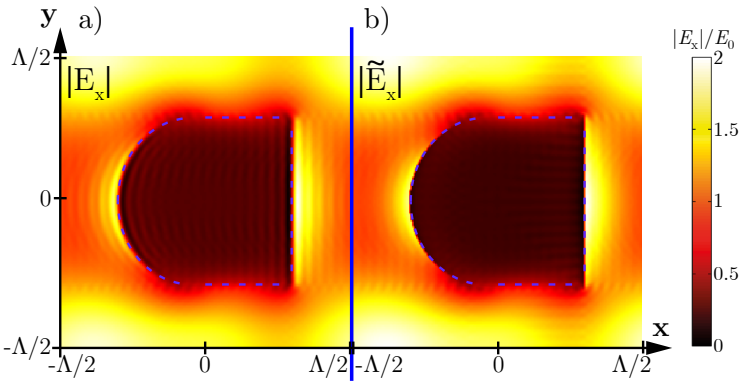

Figure 4.9: a) Spatial distribution of the dominant component of the electric field, $|E_x|$, in the 2D grating region at $z = h/2$, determined by using the conventional RCWA and $N = 27$. b) Spatial distribution of the dominant component of the electric near-field, $|\tilde{E}_x|$, determined for the same grating parameters as in a) but using the improved algorithm. The blue vertical line was added for clarity and merely separates the two plots.

plane, which was derived in Section 3.5. In each computational layer, the eigenmodes are then found numerically by solving Eq. (3.46), and the EM boundary conditions between the layers are used to determine the EM fields everywhere in the grating. In the context of the NVF formulation of RCWA for oblique 1D-periodic structures, it has been reported in Refs .[11, 22] that using an out-of-plane NVF, i.e. $\mathbf{N} = (N_x, 0, N_z)^T$, leads to an improvement of the convergence of the far-field. However, no quantitative conclusion regarding the near-field convergence of the out-of-plane NVF formulation has been made. In this section, I will study the relation between the accurate field formulation Eq. (4.9) and the near-field convergence speed for slanted 1D-periodic structures.

The situation for oblique 2D-periodic structures has not yet been explored in the context of RCWA. However, based on the results presented so far, one can conjecture that both the near- and far-field convergence of RCWA can be improved by using an out-of-plane 3D NVF, $\mathbf{N} = (N_x, N_y, N_z)^T \neq 0$, and that the near-field evaluation would be more accurate as well. The validity of this supposition will be explored in the second part of this section.

4.5.1 Analysis of slanted 1D-periodic binary diffraction gratings

The numerical characteristics of oblique 1D-periodic structures are investigated by considering the slanted binary grating depicted in Fig. 4.1(b). The period of the grating is $\Lambda = 1\,\mu\text{m}$, the filling factor, $\rho = 0.5$, the height, $h = 0.25\,\mu\text{m}$, and the slanting angle is $\gamma = \pi/4$. Only the gold grating is considered in this section as this would be the most challenging case. If the unit cell is assumed to extend from $x = -\Lambda/2$ to $x = \Lambda/2$ and the centre of the binary grating is set to be $x = 0,\ z = h/2$, a suitable out-of-plane NVF is given by

$$\tilde{N}(x, y, z) = \sqrt{2}\,\text{sign}\,(x - z)\,(1, 0, 1)^T . \tag{4.12}$$

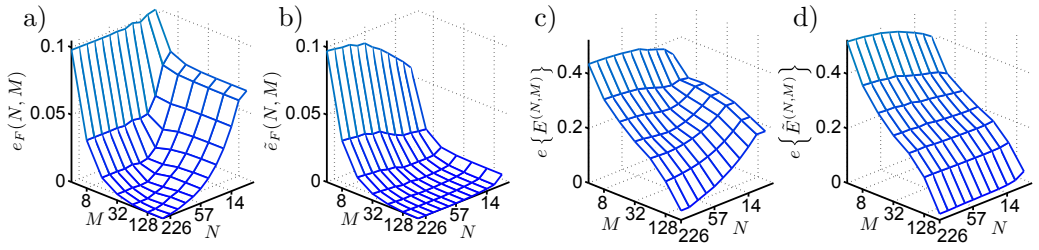

Figure 4.10: Computational results for the slanted 1D binary grating shown in Fig. 4.1(b). a) Far-field self-error $e_F(N, M)$ vs. N and M corresponding to the in-plane NVF formulation. b) Far-field self-error $\tilde{e}_F(N, M)$ corresponding to the out-of-plane NVF formulation. c) Near-field self convergence $e\{\mathbf{E}^{(N,M)}\}$ corresponding to in-plane NVF. d) Near-field self convergence $e\{\tilde{\mathbf{E}}^{(N,M)}\}$ corresponding to out-of-plane NVF.

The two formulations of the RCWA compared in this section are the in-plane NVF, $N(x, y, z) = \mathbf{e}_x$, which is used in conventional RCWA together with the conventional field evaluation Eq. (4.10), and the out-of-plane NVF, $\tilde{N}$, combined with the improved field evaluation formulation Eq. (4.9). In contrast to the results presented in the previous sections, the two formulations yield different results for both the near- and far-field quantities. For the sake of the clarity of the presentation, all physical quantities corresponding to the out-of-plane NVF formulation are denoted with a tilde symbol.

Numerical results for increasing number of harmonics, $N = 2, \ldots, 320$, and number of computational layers, $M = 2, \ldots, 256$, are presented in Fig. 4.10. It can be inferred from this figure that the in-plane formulation requires both a high number of Fourier series coefficients, N, and layers, M, to achieve convergence to a result of $R^{\text{ref}} = 0.28788$, whereas the out-of-plane formulation yields fast convergence to $\tilde{R}^{\text{ref}} = R^{(320,256)} = 0.28837$ with respect to N, as per figures 4.10(a) and (b), respectively. This behaviour is in agreement with the findings reported in Ref. [11]. The convergence of the calculated near-field, illustrated in Figs. 4.10(c) and 4.10(d), exhibits similar features. Specifically, the in-plane NVF formulation requires both high N and M to achieve a small self-error of $e\{\mathbf{E}^{(226,256)}\} = 4.7 \cdot 10^{-2}$, whereas this self-error can already be achieved with $N = 10$, $M = 256$ in the out-of-plane formulation. This clearly demonstrates a drastically improved efficiency of the near-field calculation when using the approach based on the combination of out-of-plane NVF and the accurate near-field formulation for oblique diffraction gratings. The highly improved near-field profile is illustrated in Fig. 4.11(b), which exhibits no unphysical oscillations near the gold-vacuum interface. This is in sharp contrast to the conventional field evaluation of the in-plane formulation, as per Fig. 4.11(a), which clearly suffers from spurious oscillations.

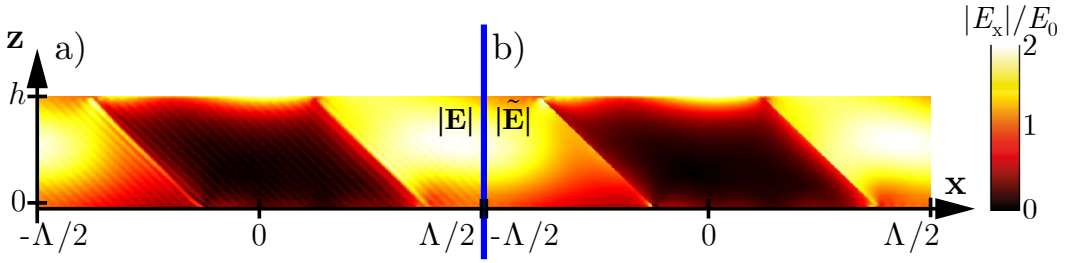

Figure 4.11: a) Spatial distribution of the electric near-field, $|E_x^{(40,256)}|$, determined using the conventional in-plane NVF. b) Spatial distribution of the electric near-field, $|\tilde{E}_x^{(40,256)}|$, determined for the same grating parameters as in a) but using the modified out-of-plane NVF formulation. The blue vertical line was added for clarity and merely separates the two plots.

4.5.2 Analysis of slanted 2D-periodic cylindrical diffraction gratings

In this section I investigate the efficiency of using the accurate field evaluation and the out-of-plane NVF formulation to model a challenging, slanted 2D-periodic diffraction grating. The grating with periods, $\Lambda_1 = \Lambda_2 = \Lambda = 1\,\mu\text{m}$, is schematically depicted in Fig. 4.1(d) and consists of a cylindrical rod with radius $r = 0.3\,\mu\text{m}$ and height $h = 0.125\,\mu\text{m}$, which is slanted by $\gamma = \pi/4$ along the x-axis. Again, only the gold grating is considered in this section. Finally, the incident plane wave is impinging onto the grating along the normal direction, is polarised along the x-axis, and has a wavelength of $\lambda = 2\,\mu\text{m}$.

The reflection coefficient calculated for $N = 3, \ldots, 19$ harmonics and $M = 2, \ldots, 32$ layers is shown in Fig. 4.12(a) and Fig. 4.12(b) and was determined by using the conventional in-plane NVF and the out-of-plane NVF formulation, respectively. It can be seen that both approaches converge rather slowly, neither one achieves convergence even for the highest considered values of $M = 32$ and $N = 19$. Moreover, in order to characterise the error of the near-field calculations in the two formulations, the self-error with respect to the reference solutions obtained with $N = 19$ and $M = 32$ is presented in figures 4.12(c) (d). The in-plane formulation achieves a relative self-error of $e\{\mathbf{E}^{(13,32)}\} = 0.625$, whereas for the same values of N and M, the modified field evaluation in conjunction with the out-of-plane NVF achieves a substantially lower self-error of $e\{\mathbf{E}^{(13,32)}\} = 0.261$. It has to be stressed, that the necessary accuracy for full convergence could not be achieved in these simulations. The evolution of the computational results for the shown values of $N \leq 19$ and $M \leq 32$ however can be interpreted in favour of the out-of-plane NVF formulation, due to the lower near-field self-error $e\{\mathbf{E}^{(13,32)}\} < e\{\mathbf{E}^{(13,32)}\}$. It can be supposed that future simulations with finer discretisation will reveal the practical benefit of the out-of-plane NVF formulation in conjunction with the modified field formulation for 2D-periodic slanted structures, similar to the case of 1D-periodic slanted diffraction gratings.

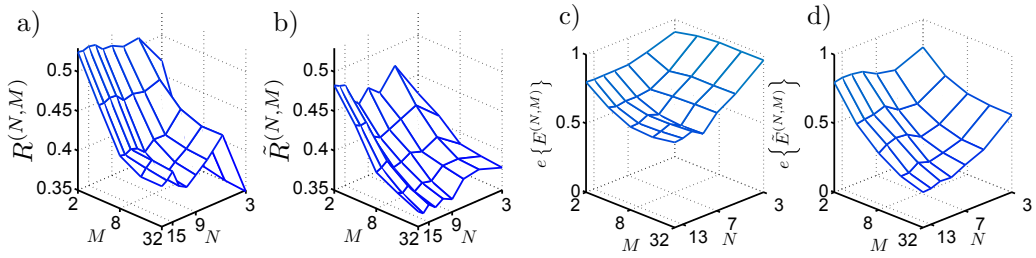

Figure 4.12: Computational results for the slanted 2D cylindrical grating shown in Fig. 4.1(d). a) Reflection coefficient R vs. N and M, determined using the in-plane NVF formulation. b) Reflection coefficient $\tilde{R}$ vs. N and M, determined using the out-of-plane NVF formulation. c), d) Near-field convergence of conventional and modified formulation, respectively.

4.6 Concluding remarks on accurate near-field evaluation in RCWA

To summarise, I have analysed the numerical near-fields calculated using the RCWA and identified the Gibbs phenomenon as the main reason for their slow convergence and the spurious oscillations from which they suffer. As a solution to these deficiencies of RCWA, I proposed a modification of this method that can be applied for modelling arbitrary diffraction gratings and, more generally, periodic optical structures. My modified formulation significantly improves the accuracy of 1D-RCWA calculations at negligible additional computational cost for both straight and slanted gratings, where it speeds up convergence and effectively removes the numerical artefacts from the calculated near-fields. The accuracy of 2D-periodic grating simulations can be enhanced and the malicious effect of the Gibbs phenomenon can be reduced as well. However, the advantage here is less pronounced than in the 1D-periodic case. The reduced performance in 2D can be attributed the fact that usually no exact formula for the NVF, which is at the core of the modified formulation, is available, and that is often suffers from discontinuities. Therefore, it might be fruitful to investigate more elaborate NVF-formulations and their suitability for near-field calculations, such as a complex valued NVF [26], which is continuous everywhere in the grating region.

It can be expected that my proposed modification of the RCWA method will greatly advance its capabilities, especially for 1D-periodic optical structures. In particular, this improved method could prove instrumental to accurate modelling of plasmonic periodic structures, diffraction gratings, and surface-nonlinear devices, namely to simulation of physical systems whose functionality rely on the electromagnetic near-field at interfaces. And in fact, the accurate near-field evaluation will be a crucial ingredient in the accurate numerical modelling of nonlinear 2D materials in diffraction gratings. Before investigating this in Chapter 6, the details of the implementation of the RCWA in in a commercially available computer code will be discussed in the next chapter.

Bibliography

[1] M. G. Moharam and T. K. Gaylord, "Rigorous coupled-wave analysis of planar-grating diffraction," J. Opt. Soc. Am. **71**, 811–818 (1981).

[2] M. G. Moharam and T. K. Gaylord, "Rigorous coupled-wave analysis of metallic surface-relief gratings," J. Opt. Soc. Am. A. **3**, 1780–1787 (1986).

[3] M. G. Moharam, T. K. Gaylord, 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).

[4] P. Lalanne and G. M. Morris, "Highly improved convergence of the coupled-wave method for tm polarization," J. Opt. Soc. Am. A. **13**, 779–784 (1996).

[5] L. Li, "Use of fourier series in the analysis of discontinuous periodic structures," J. Opt. Soc. Am. A. **13**, 1870–1876 (1996).

[6] L. Li, "New formulation of the fourier modal method for crossed surface-relief gratings," J. Opt. Soc. Am. A. **14**, 2758–2767 (1997).

[7] S. Essig and K. Busch, "Generation of adaptive coordinates and their use in the fourier modal method," Opt. Express **18**, 23258–23274 (2010).

[8] 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**, 2880–2890 (2007).

[9] L. Li, "Fourier modal method for crossed anisotropic gratings with arbitrary permittivity and permeability tensors," J. Opt. Soc. Am. A. **5**, 345 (2003).

[10] M. G. Moharam, D. A. Pommet, E. B. Grann, and T. K. Gaylord, "Stable implementation of the rigorous coupled-wave analysis for surface-relief gratings: enhanced transmittance matrix approach," J. Opt. Soc. Am. A. **12**, 1077–1086 (1995).

[11] I. Gushchin and A. V. Tishchenko, "Fourier modal method for relief gratings with oblique boundary conditions," J. Opt. Soc. Am. A. **27**, 1575–1583 (2010).

[12] K. Kneipp, Y. Wang, H. Kneipp, L. T. Perelman, I. Itzkan, R. R., Dasari, and M. S. Feld, "Single molecule detection using surface-enhanced raman scattering (sers)," Phys. Rev. Lett. **78**, 1667 (1997).

[13] J. A. H. Van Nieuwstadt, M Sandtke, R. H. Harmsen, F. B. Segerink, J. C. Prangsma, S. Enoch, and L Kuipers, "Strong modification of the nonlinear optical response of metallic subwavelength hole arrays," Phys. Rev. Lett. **97**, 146102 (2006).

[14] L. Cao, N. C. Panoiu, and R. M. Osgood, "Surface second-harmonic generation from surface plasmon waves scattered by metallic nanostructures," Phys. Rev. B **75**, 205401 (2007).

[15] C. G. Biris and N. C. Panoiu, "Second harmonic generation in metamaterials based on homogeneous centrosymmetric nanowires," Phys. Rev. B **81**, 195102 (2010).

[16] J. Homola, S. S. Yee, and G. Gauglitz, "Surface plasmon resonance sensors: review," Sens. Act. B: Chem. **54**, 3–15 (1999).

[17] A. V. Kabashin, P. Evans, S. Pastkovsky, W. Hendren, G. A. Wurtz, R. Atkinson, R. Pollard, V. A. Podolskiy, and A. Zayats, "Plasmonic nanorod metamaterials for biosensing," Nat. Mater. **8**, 867–871 (2009).

[18] N. Liu, M. Mesch, T. Weiss, M. Hentschel, and H. Giessen, "Infrared perfect absorber and its application as plasmonic sensor," Nano Lett. **10**, 2342–2348 (2010).

[19] C. G. Biris and N. C. Panoiu, "Excitation of dark plasmonic cavity modes via nonlinearly induced dipoles: applications to near-infrared plasmonic sensing," Nanotechnol. **22**, 235502 (2011).

[20] A. J. L. Adam, J. M. Brok, M. A. Seo, K. J. Ahn, D. S. Kim, J. H. Kang, Q. H. Park, M. Nagel, and P. C. M. Planken, "Advanced terahertz electric near-fieldmeasurements at sub-wavelength diameter metallic apertures," Opt. Express **16**, 7407–7417 (2008).

[21] P. Lalanne and M. P. Jurek, "Computation of the near-field pattern with the coupled-wave method for transverse magnetic polarization," J. Mod. Opt. **45**, 1357–1374 (1998).

[22] E. Popov, M. Neviere, B. Gralak, and G. Tayeb, "Staircase approximation validity for arbitrary-shaped gratings," J. Opt. Soc. Am. A. **19**, 33–42 (2002).

[23] M. Weismann, D. F. G. Gallagher, and N. C. Panoiu, "Accurate near-field evaluation in the rigorous coupled-wave analysis," J. Opt. **17**, 125612 (2015).

[24] N. M. Lyndin, O. Parriaux, and A. V. Tishchenko, "Modal analysis and suppression of the fourier modal method instabilities in highly conductive gratings," J. Opt. Soc. Am. A. **24**, 3781–3788 (2007).

[25] P. B. Johnson and R. W. Christy, "Optical constants of the noble metals," Phys. Rev. B **6**, 4370–4379 (1972).

[26] 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**, 27511–27524 (2010).

Chapter 5

Software implementation details of the rigorous coupled-wave analysis: OmniSim RCWA

5.1 Introduction

This chapter discusses the computer implementation of the rigorous coupled-wave analysis in a reliable, efficient and user-friendly software tool named OmniSim RCWA [1]. OmniSim RCWA follows the mathematical formulation of Chapter 3 and includes the accurate near-field calculation method from Chapter 4. Therefore, it is a general tool for the accurate simulation of periodic structures, including diffraction gratings, diffractive optical elements and periodic metamaterials.

As mentioned earlier, RCWA is a widely used algorithm for modelling 1D- and 2D-periodic diffraction gratings. Many researchers develop and maintain their own hand-tailored RCWA software implementations, especially for 1D-periodic structures where the Fourier factorisation problem can be solved easily, see Section 3.3.4. In addition, there is a number of commercially [2] and freely available implementations of RCWA [3]. However, free implementations and academic implementations often require expert knowledge of the underlying numerical method in order to use them. They are often poorly or not at all documented, they are typically not regularly maintained, and there is no technical support available to help with user problems. Moreover, the specification of grating structures in such implementations is typically cumbersome as they lack the integration into computer aided design (CAD) software. OmniSim RCWA, on the other hand, is different from freely available and academic codes and overcomes their shortcomings by integrating the RCWA algorithm into a mature CAD software, which greatly simplifies specification of arbitrarily complex structures, eases the setup of simulations and allows detailed post-processing of results. To this end, OmniSim RCWA was conceived in a collaboration with the company Photon Design Ltd, and is now part of their software suite. Photon Design specialises

in high quality simulation software for rigorous modelling of photonic structures, devices and circuits. Their software repertoire contains a 2D and 3D optical waveguide modelling software, a photonic crystal design program, a CAD-integrated FDTD/FETD solver, an active component design software and a photonic circuit and laser diode simulator. This is accompanied by an automated optimization toolbox and a powerful CAD editor.

Software development in a commercial environment has different requirements from those in an academic environment: *i)* The user is not the developer. Therefore a clear manual and instructive examples must be provided. *ii)* The implementation is not a standalone program but is integrated into a larger software suite. Hence, the implementation has to be well documented and adhere to specific guidelines. This is related to the fact that such an implementation is likely to be maintained over a long period of time and by different developers who are not familiar with all aspects of the underlying numerical method or its software implementation. *iii)* Reliability, error handling and validation: academic software is often used to validate theoretical concepts by means of specific examples. Commercial software, by contrast, has to work efficiently for any user-defined input structure, or at least clearly define the range of parameters for which it works. Moreover, errors must be detected and handled automatically and reliably. And although validation is important for any software implementation, special attention was paid to benchmarking and validation of OmniSim RCWA by comparison of simulation results with analytical or experimental results, with implementations of different numerical methods or with an independent implementation of RCWA.

The remainder of this chapter will explain the implementation of OmniSim RCWA and discuss its characteristics. Section 5.2 will explain the program flow of OmniSim RCWA and present how the software implements the mathematical formulation of RCWA. The specific algorithms developed for OmniSim RCWA will be described and the computational complexity of RCWA is discussed in Section 5.3, before final conclusions about this highly sophisticated software tool are drawn in Section 5.4.

5.2 Program flow of OmniSim RCWA

OmniSim RCWA is a complex software tool, because it allows the efficient simulation of different kinds of diffraction gratings, such as multilayer, slanted, holographic, and 1D- and 2D-periodic gratings made of dielectrics, semiconductors or metals. The formulation of RCWA is general and allows the modelling of arbitrary diffraction grating structures. However, only the adequate choice of algorithms and numerical procedures to implement the mathematical formulation will guarantee the high efficiency and accuracy of simulations in OmniSim RCWA. This section describes how OmniSim RCWA implements the mathematical steps discussed in the previous chapters and how it makes them accessible to the user of the software.

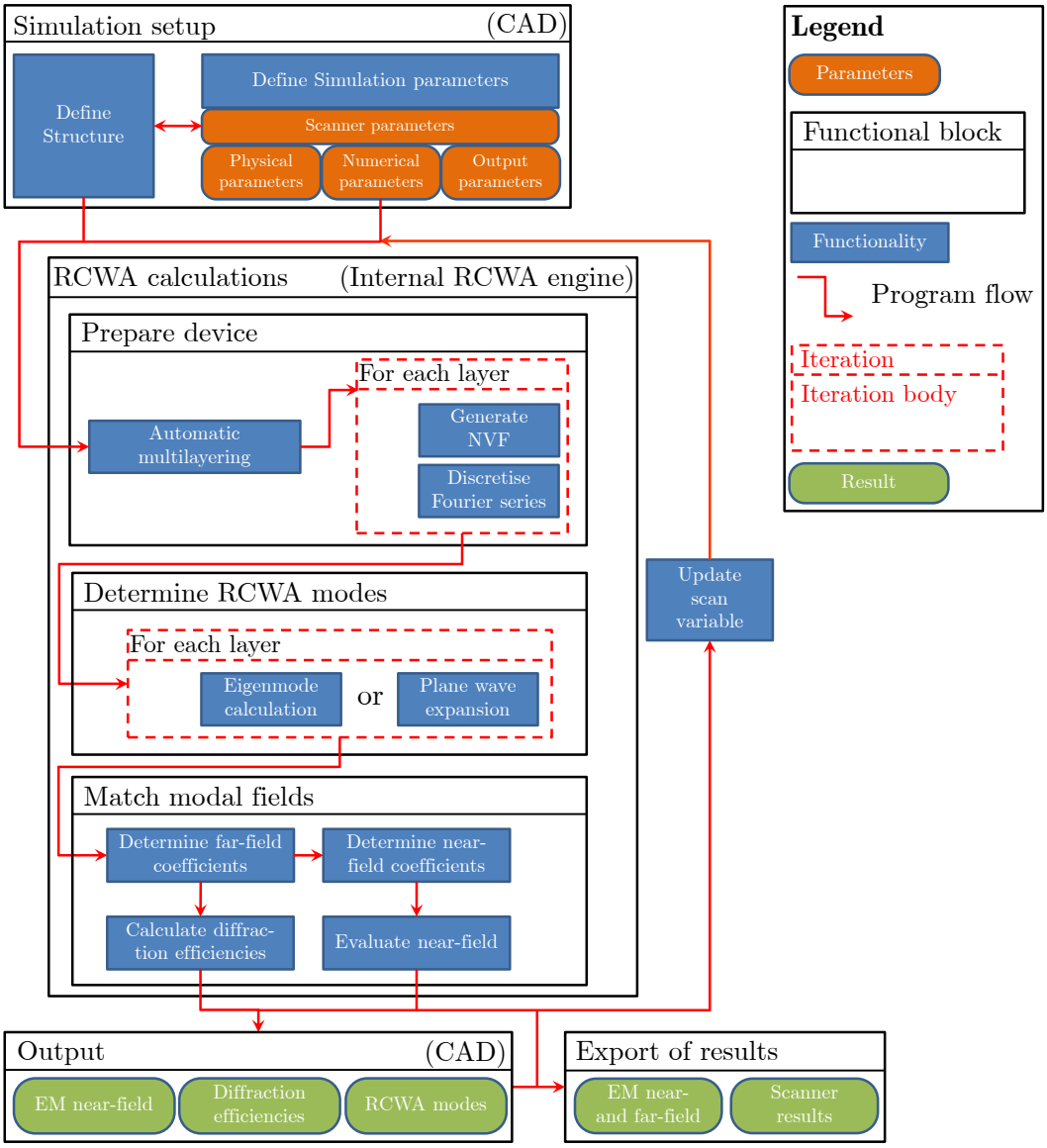

Figure 5.1: Work flow of OmniSim RCWA.

Each OmniSim RCWA simulation has three stages and follows a typical work flow depicted in Fig. 5.1. Only the first and third stages, the simulation setup and the output and export, respectively, are accessible to the user. The most important stage, the actual RCWA calculations, are handled internally by the numerical engine and are hence shielded from the user. Before going into detail, it should be briefly mentioned that OmniSim RCWA has two interactive modes of operation, the first being a single simulation for a specific set of parameters, whereas the other mode is a one-dimensional scan of any variable, e.g. wavelength of excitation, polarisation angle or a geometric quantity that characterises the device at hand.

The first stage of a simulation is the *setup of the computational process* and has two

parts: the user defines the diffraction grating structure in the CAD layout editor with the help of predefined components or manually defined shapes. Their optical properties can be specified manually or by means of a large library of materials. The other part of the simulation setup is the definition of the simulation parameters. They are of three kinds: the physical simulation parameters are the incident angles, polarization stage and wavelength of the simulation. The numerical parameters are the number of harmonics, N, to be used in the simulation, the number of layers N_L used in the automatic multi-layering as well as algorithmic choices for the Fourier factorisation (*zig-zag* approximation and NVF factorisation). The third kind of parameters determines the output of the simulation. Especially when simulating 2D-periodic multilayered structures, it is impractical to output all computed physical quantities. Instead, certain diffraction efficiencies can be selected to characterise the far-field, and a 3D grid to evaluate the near-field can be specified.

A special role amongst the parameters is played by the scanner parameters. OmniSim RCWA can automatically perform a study of a finite range of virtually any scalar quantity. This can not only be any of the parameters above (e.g. wavelength, number of harmonics N, or 3D output grid size), but in fact any variable used to specify the grating structure in the CAD.

Having defined all simulation parameters, the next stage is entered, namely the actual *RCWA calculation*. The RCWA calculations have three steps, which are entirely automated and are performed by the *numerical RCWA engine*. The first step samples the device that was specified in the CAD and discretises it appropriately for the RCWA calculations. In the case when the structure under investigation is not homogeneous in the longitudinal direction, i.e. if it is a multilayer or slanted grating, the N_L computational layers are defined automatically. For each of these layers, one or two tasks are performed by the RCWA engine: If the *zig-zag* approximation is used, the Fourier series coefficient matrices $\lfloor\lceil\epsilon_r\rceil\rfloor$ and $\lceil\lfloor\epsilon_r\rfloor\rceil$ are calculated according to Eq. (3.36). If the NVF factorisation is employed, the NVF, $\mathbf{N}$, of each layer is generated according to Section 5.3.2 and their vectors of Fourier series coefficients, $[N_\alpha N_\beta]$, are determined by numerical integration. Independent of the factorisation rule, the vectors of Fourier series coefficients $[\epsilon_r]$ and $[1/\epsilon_r]$ for each layer are calculated according to Eq. (3.19). All numerical integrations are not directly performed by evaluating their defining integrals, e.g. Eq. (3.19), but instead they are rewritten and calculated by the fast Fourier transform (FFT) with a high number of sampling points to ensure that no integration error enters the Fourier series discretisation of the structure in the transverse plane, see Appendix C.1. Upon the completion of this process, the simulation device is fully discretised in N_L layers, each of which being discretised by the structural Fourier series coefficients.

In the second step, the modal representation of the EM fields Eq. (3.55) therein are

determined by solving numerically Eq. (3.46), or their respective simplifications for 1D-periodic gratings. In the case when a computational layer is homogeneous, the modes are directly given by the plane wave expansion presented in Appendix B.

The third and final step of the RCWA calculation is the matching of the modal fields in different layers. The far-field coefficients are determined from Eq. (3.69) and the diffraction efficiencies are evaluated by Eq. (3.70). If the near-field is requested, all modal coefficients are determined efficiently following Section 5.3.3. The near-field is evaluated according to Eq. (4.10) or Eq. (4.9) in case the NVF factorisation is employed. If specified, the EM fields are directly exported to the hard-drive for post processing. In the case when a single simulation has been specified, OmniSim RCWA leaves the numerical RCWA engine and enters the last stage, namely the inspection of the output. Otherwise the scan variable is updated and another RCWA calculation is performed.

The *output* of an RCWA calculation is presented interactively in three complementary ways. The EM near-fields can be inspected in the CAD editor, scalar output quantities are displayed and – in the case of a scan – plotted, and the discretisation and the RCWA modes of each layer can be inspected in the OmniSim RCWA Mode Explorer to gain additional insight into the diffraction grating at hand. All output quantities can be exported for further investigation by using OmniSim RCWA or other tools such as MATLAB or Python.

5.3 Algorithmic choices in OmniSim RCWA

The description in the last section focussed on the software implementation of OmniSim RCWA and how it relates to the mathematical formulation. In this section the specific algorithms of the numerical engine are discussed.

5.3.1 Numerical linear algebra

The efficiency of OmniSim RCWA is strongly linked to the choice of numerical routines to handle matrix algebraic operations prevalent in RCWA. Therefore, functions from Photon Design's proprietary, highly optimized mathematical library were used to implement the solution of various mathematical operations. They include the solution of the complex, dense, and potentially large linear eigenvalue problems, Eq. (3.46), and the matrix multiplications and inversions employed throughout the $\mathcal{S}$-matrix calculations Eq. (3.67). Moreover, the FFT are implemented in this library, which is used in the discretisation of the grating structure Eq. (3.19) and in the evaluation of the EM fields Eq. (4.9). The benefits of using such a library are apparent: it is highly optimized, validated, and regularly maintained and updated. Therefore, improvements and updates to the numerical routines in Photon Design's mathematical library will directly carry over to OmniSim RCWA. Apart from that, special routines to handle Toeplitz matrices in a memory-saving and computationally efficient way were implemented according to Appendix C.

The two remaining specific algorithms to automatically generate a NVF and to efficiently calculate the modal coefficients in each computational layer are presented now.

5.3.2 Automated generation of normal vector fields

The NVF factorisation approach yields an accurate formulation of the RCWA, which allows fast-converging solutions of RCWA simulations, as was demonstrated in Section 3.6. The automated generation of a NVF has, however, not been addressed here. In the context of RCWA, analytical formulae or semi-analytical algorithms for N_α are used [4], but this is not an option for OmniSim RCWA, as they are not available for generic structures a user might define. Different algorithms for the generation of NVFs have been proposed in the context of RCWA, such as optimized inverted distance weighting [5], or have been borrowed from computer graphics [3, 6].

Such an algorithm has to fulfil several criteria. It has to yield accurate normals at material interfaces and a smooth continuation into the unit cell with a small number of discontinuities away from the material interfaces. Moreover, the resulting NVFs need to be periodic and need to transform under affine operations as the structure does. Furthermore, the symmetry properties of a device need to be preserved in the generated NVF. Finally, the algorithm has to have small computational cost to not considerably increase the overall runtime of an RCWA software implementation.

The main steps of the NVF generation in OmniSim RCWA are:

1. Discretise the cross section of material parameters in the unit cell at a set of grid points.
2. Determine the discrete material boundary points.
3. Calculate the normals at all boundary points.
4. Iteratively propagate the known normal values to their neighbouring grid points by a suitable averaging procedure.

Theoretical considerations and numerical tests showed that this algorithm fulfils the requirements for NVF generation in the context of RCWA.

5.3.3 Fast calculation of modal coefficients

The calculation of the modal field coefficients in a multilayer structure is given in Eq. (3.63). However, the direct implementation of this in a multilayer structure with N_L layers is inefficient as its computational cost is of the order $\mathcal{O}(N_L^2 N_0^3)$, where the factor N_0^3 is due to the computational cost of the calculation and concatenation of $\mathcal{S}$-matrices of size $4N_0 \times 4N_0$. A more efficient approach to near-field evaluation is taken in OmniSim RCWA and this will be described now.

To this end, consider the schematic of a multilayer structure in Fig. 5.2(a) where it is assumed that the modal information, i.e. the mode profiles $\mathsf{E}_l^\pm$ and $\mathsf{H}_l^\pm$ and the mode

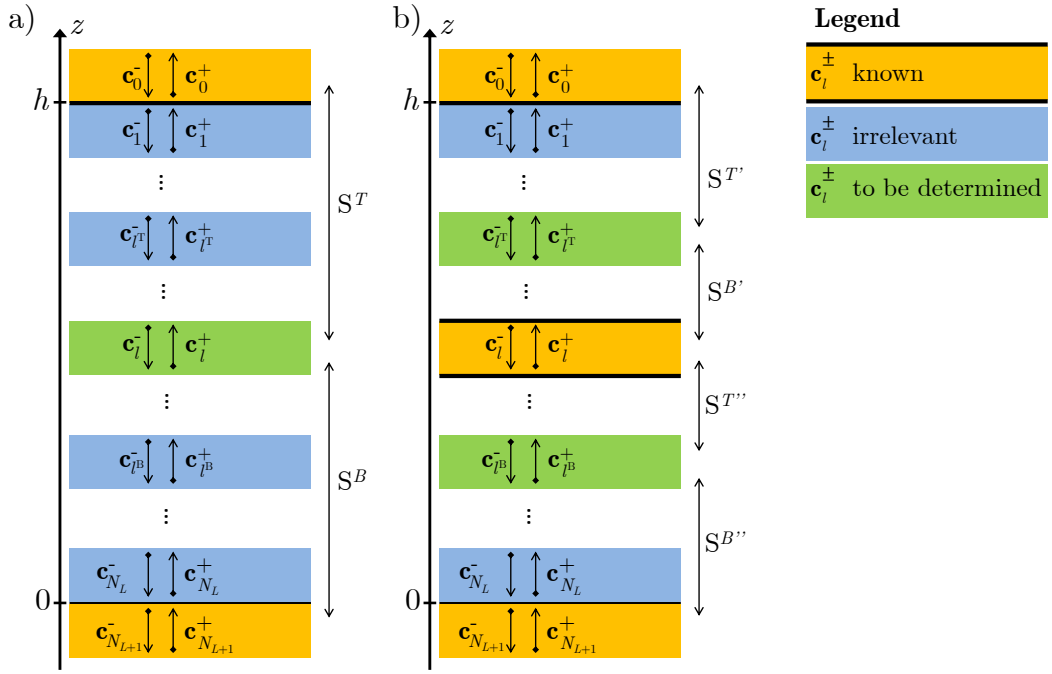

Figure 5.2: Recursive calculation of mode coefficients in OmniSim RCWA. a) Schematic representation of a N_L-layer structure with known cover and substrate coefficients $\mathbf{c}_0^{\pm}$ and $\mathbf{c}_{N_L+1}^{\pm}$. The coefficients in layer $l = \lfloor N_L/2 \rfloor$ are calculated in the first step. b) In the second step, the coefficients in layer l are known and the subinterval of layers between cover and layer l around l^T resembles the initial situation in (a), where layer l takes the role of the known bottom layer.

propagation constants $\mathsf{V}_l^{\pm}$ are given in each layer, $l = 0, \ldots, N_{L+1}$. Moreover, the incident and outgoing far-field coefficients are given and determined by Eq. (3.69), respectively. The calculation of the mode coefficients is performed recursively. In the first step, $n = 1$, the coefficients $\mathbf{c}_{\lfloor N_L/2 \rfloor}^{\pm}$ in the middle layer are calculated by setting up the $\mathcal{S}$-matrices of all top and bottom layers, S^T and S^B, respectively, and applying Eq. (3.63) for $\mathrm{S}^{AB} = \mathrm{S}^T$ and $\mathrm{S}^{BC} = S^B$. The total number of $\mathcal{S}$-matrix products Eq. (3.66) is $\mathcal{O}(2 \cdot N_L/2)$. Since the coefficients in layer $\lfloor N_L/2 \rfloor$ are known, one can identify the remaining layers below $l = \lfloor N_L/2 \rfloor$ with the initial setting, and similarly the remaining top layers, as per Fig. 5.2(b).

This procedure is repeated until a subinterval of layers is empty, which will occur after a maximum of $\lceil \log_2(N_L) \rceil$ recursions. Because the nth recursion handles 2^{n-1} subintervals of length $N/2^n$, a total of $\mathcal{O}(N_L \log(N_L))$ $\mathcal{S}$-matrix operations are necessary. Therefore, the computational cost for the evaluation of mode coefficients in OmniSim RCWA is of the order $\mathcal{O}(N_L \log N_L N_0^3)$.

5.3.4 Parallelisation of OmniSim RCWA

OmniSim RCWA is shared memory parallel on two levels of granularity, which are handled by different systems of parallelisation. The outer level of parallelism is at the level of the computational layers and is handled by Photon Design's proprietary parallel library.

Referring back to Fig. 5.1, one finds that the generation of NVFs, the discretisation of the structure in the transverse direction, and the determination of the RCWA modes in one layer are independent of the state of the calculation in other layers. They can be done in parallel. Apart from that, the recursive calculation of modal coefficients in Section 5.3.3 is performed parallel by OmniSim RCWA, and the evaluation of the EM near-field in the grating structure at different z is inherently parallel. Especially in the case of gratings with a small number of computational layers the outer level of parallelisation just described does not take full advantage of modern multi-core computer systems. Therefore, OmniSim RCWA also employs parallel implementation of basic linear algebraic routines as an inner level of parallelism.

5.3.5 Remarks on computational complexity

Having described the algorithms in the implementation of OmniSim RCWA, one can see that its highest computational cost is the evaluation of the EM field coefficients in the grating structure, which has an asymptotic runtime of $\mathcal{O}(N_L \log N_L N_0^3)$, even in the improved version of Section 5.3.3. This is followed by the calculation of the far-field coefficients and the solution of N_L eigenvalue problems in the computational layers, which each requires $\mathcal{O}(N_L N_0^3)$ operations. This dominates by far the computational cost of all pre-processing steps, namely the NVF generation in Section 5.3.2 and the numerical integration Eq. (3.19) in the definition of $[\epsilon_r]$, for any practical choice of sampling points.

An additional concern when analysing the computational complexity is memory consumption. The most demanding part of RCWA is the representation of the eigenmodes in each of N_L computational layers by $\mathsf{E}_l^{\pm}$ and $\mathsf{H}_l^{\pm}$ in Eq. (3.55) with a total memory consumption of $\mathcal{O}(N_L N_0^2)$. If only the cover and substrate fields are required as simulation results, this is reduced to $\mathcal{O}(N_0^2)$ because only the mode-information of two subsequent computational layers is simultaneously required to realize Eq. (3.69).

These theoretical considerations show the scaling of RCWA, namely the cubic and quadratic dependence of runtime and memory consumption, respectively, on the number of orders N_0. Whereas the simulation of 1D periodic structures with hundreds of diffraction orders, $N_0 = (2N_1 + 1)$, and layers can be handled by modest computer hardware, the simulation of 2D periodic structures, where $N_0 = (2N_1+1)(2N_2+1)$ is limited to moderate numbers of harmonics. This is why correct Fourier factorisation from Section 3.3.3 is of such importance to the efficiency of RCWA, because fast Fourier factorisation drastically reduces the number of orders necessary to achieve converge of RCWA simulations even for complex 2D-periodic structures.

5.4 Conclusions

This chapter discussed the computer implementation of the RCWA method in an efficient, versatile and user friendly software tool, OmniSim RCWA, for the investigation of arbitrary diffraction grating structures. OmniSim RCWA is the first, and to the best of my knowledge, the only implementation of the RCWA with accurate near-field evaluation for 2D periodic structures and oblique structures, and the only implementation of RCWA with NVF factorisation and automatic NVF generation which is fully integrated into a powerful CAD editor. Moreover, it makes use of current parallelisation techniques to take full advantage of modern multi-core computer systems.

Bibliography

[1] Photon Design Ltd., *OmniSim/RCWA*, `http://www.photond.com/products/omnisim.htm`.

[2] RSoft Design Group, *DiffractMOD*, `http://www.rsoftdesign.com/`.

[3] V. Liu and S. Fan, "S 4: a free electromagnetic solver for layered periodic structures," Computer Physics Communications **183**, 2233–2244 (2012).

[4] 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**, 2880–2890 (2007).

[5] 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**, 17295–17301 (2008).

[6] M. Fisher, P. Schröder, M. Desbrun, and H. Hoppe, "Design of Tangent Vector Fields," ACM Trans. Graph. **26** (2007).

Chapter 6

Theoretical and computational analysis of second- and third-harmonic generation in periodically patterned 2D-3D heteromaterials

6.1 Introduction

The advent of novel two-dimensional materials, most notably graphene [1–3], has created a large demand for the accurate and efficient numerical simulation of periodic structures comprising both regular, three-dimensional, materials and unconventional 2D materials.

As was described in Section 2.5, there is a multitude of numerical methods for computational study of conventional optical structures and devices comprising 3D materials and they can in principle be used for modelling 2D materials (2DM), too. This is customarily done by defining an effective thickness of the material and incorporating the 2DM into the computational algorithm simply as a very thin layer of 3D material [4–6]. This approach, albeit simple and easy to implement, has a serious drawback, namely it relies on an obviously ambiguous quantity, the thickness of the monolayer. Moreover, this thickness is typically orders of magnitude smaller than the other characteristic lengths of the photonic structure and the operating wavelength, leading to a large length-scale imbalance that is detrimental to the effectiveness of the spatial discretisation. These issues result in potentially reduced accuracy, increased computational cost, and numerical artefacts that are difficult to avoid [7, 8]. It is hence desirable to treat the sources of the optical effects pertaining to the 2DM as confined to a 2D manifold, i.e. as induced by a surface conductivity [7, 9, 10].

The numerical method proposed in this chapter follows this approach and implements it in the context of the RCWA method by building on the formulation from Chapters 3 and 4. However, in addition to previous work on handling 2DMs by means of Fourier series expansion methods [8, 11], the method presented here also models nonlinear optical effects of

2D materials, in particular third- and second-harmonic generation in graphene and TMDC monolayer materials, respectively, and provides the mathematical formulation for arbitrary nonlinear processes that can be treated in the undepleted pump approximation. Additionally, the method introduced in this work allows one to describe structured 2D materials, which themselves can be embedded in inhomogeneous 3D structures.

The remainder of this chapter introduces the general periodic optical system under consideration in Section 6.2 by extending the setting from Section 3.2. A detailed overview of linear and nonlinear optical properties of different 2D materials is given in Section 6.3. Section 6.4 provides the mathematical formulation for harmonic generation in patterned 2D materials based on the RCWA and introduces the inhomogeneous scattering matrix formalism for multilayered 3D structures in Section 6.4.4. This is followed by Section 6.5, where the validation and benchmarking of the numerical method is performed, using as test problems 1D- and 2D-periodic structures. Applications of the proposed numerical method are presented and discussed in Section 6.6, where different resonant mechanisms to enhance the nonlinear optical response of 2D-3D heteromaterials containing TDMC monolayers and nanostructured graphene are investigated. Finally, the main conclusions are drawn and an outlook towards future work and the potential of the proposed numerical method is given in Section 6.7.

6.2 Photonic system: Geometry and materials parameters

The computational method introduced in this chapter is designed for a very general physical setting, namely periodically patterned photonic structures that contain both bulk and 2D optical materials. This general photonic system is schematically illustrated in Fig. 6.1, which now incorporates periodically patterned sheets of 2D materials into the grating structure presented in Section 3.2. These additional 2DM sheets, or simply sheets, are located at $z = z_s$ and are assumed to lay in the (x, y)-plane between two bulk-layers. Each sheet is made of homogeneous or periodically patterned 2DMs and is described by its surface conductivity distribution, $\sigma_s(x, y, z_s)$. Equally important, since the nonlinear optical response of different 2D materials contained by the photonic structure is described via a generic nonlinear polarisation, this computational method can be used to investigate a multitude of nonlinear optical effects, including SHG, sum- and difference-frequency generation, THG, and four-wave mixing.

Before the RCWA for linear and nonlinear light interaction with 2D materials is derived thoroughly in Section 6.4, the optical properties of 2DMs are introduced and formalised mathematically in the subsequent section.

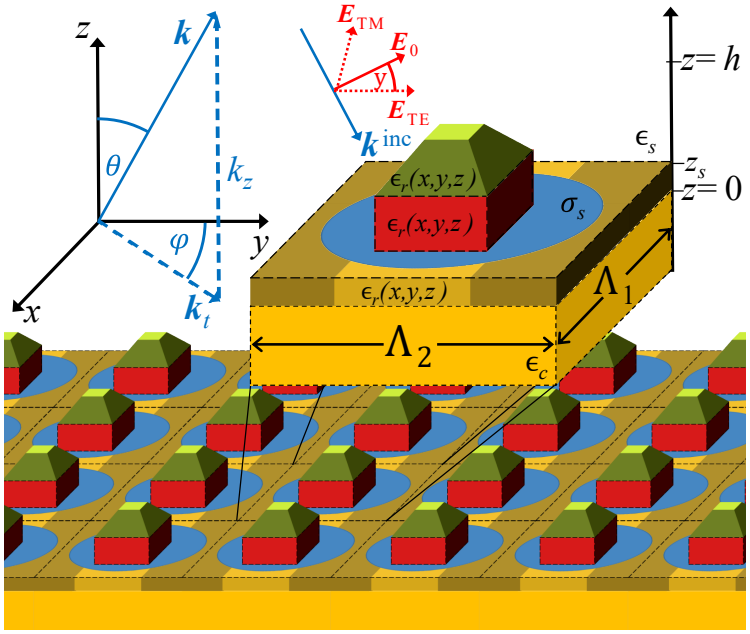

Figure 6.1: Schematic of a generic multilayer x, y-periodic structure under plane wave incidence and a close-up of one unit cell. The distribution of the 2D material (blue) is defined by the position dependent sheet conductance $\sigma_s(x, y, z_s)$ at $z = z_s$. The bulk part of the periodic structure, which can consist of slanted (olive green), straight walled (red) or embedded parts (sand brown), is defined by the permittivity function $\epsilon_r(x, y, z)$.

6.3 Linear and nonlinear optical properties of 2D materials

Understanding the physics of 2DMs is a rather recent endeavour and as such a comprehensive characterisation of their linear and nonlinear optical properties is only emerging. In particular, a complete knowledge of the frequency dependence of the linear and nonlinear optical constants is a prerequisite for a rigorous computational description of the optical response of these materials. Theoretical calculations [12–15] and experimental investigations [16–18] are beginning to fill in this critical gap, a summary of the results of these studies being briefly outlined in this section.

6.3.1 Linear optical properties of 2D materials

Since graphene and other 2DMs are physical systems consisting of a single atomic layer, their electromagnetic properties are conveniently characterised by surface quantities. For example, in the case of graphene, the optical properties are described by the sheet conductance (sometimes simply called conductivity), which in the random phase approximation and at zero temperature can be expressed as [12, 19]:

$$\frac{\sigma_s(\omega)}{\sigma_0} = \frac{4\varepsilon_F}{\pi\hbar}\frac{i}{\omega + i\tau^{-1}} + \theta(\hbar\omega - 2\varepsilon_F) + \frac{i}{\pi}\ln\left|\frac{\hbar\omega - 2\varepsilon_F}{\hbar\omega + 2\varepsilon_F}\right|. \tag{6.1}$$

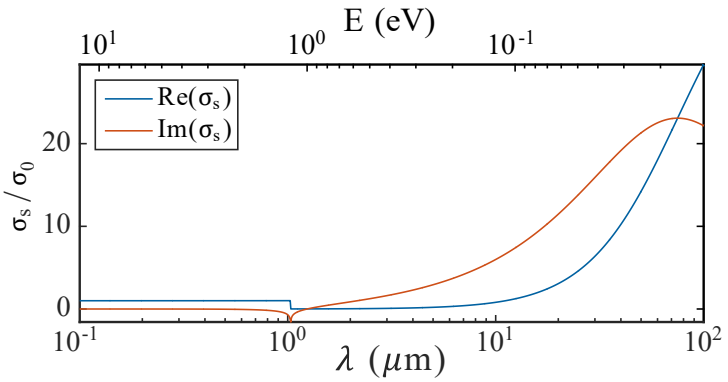

Figure 6.2: Dispersion of the complex sheet conductance $\sigma_s(\omega)$ of graphene in the interband and intraband wavelength range.

Here, $\sigma_0 = e^2/(4\hbar) = 6.0853 \times 10^{-5}\,\mathrm{A\,V^{-1}}$ is the universal dynamic conductivity of graphene, e denotes the elementary charge, ε_F the Fermi level, $\hbar$ the reduced Planck constant, τ the relaxation time and $\theta(\cdot)$ is the Heaviside step function. The values for the Fermi level and relaxation time used throughout this chapter are $\varepsilon_F = 0.6\,\mathrm{eV}$ and $\tau = 0.25\,\mathrm{ps}/2\pi$, respectively. The dispersion of $\sigma_s(\omega)$ for graphene is shown in Fig. 6.2.

The sheet conductance contains the intraband (Drude) contribution, described by the first term in the r.h.s. of Eq. (6.1), and the interband component given by the next two terms. A more complete model valid at finite temperature can be used (see Refs. [19, 20]); however, the model described by Eq. (6.1) already captures the main features of the graphene's conductivity, most notable being the possibility of tuning $\sigma_s(\omega)$ by changing the Fermi level via chemical doping or applying a gate voltage.

In many numerical methods pertaining to computational electromagnetics it is more convenient to work with bulk rather than surface quantities and therefore one often introduces bulk equivalents of the surface quantities. In particular, instead of the sheet conductance, σ_s, one uses a bulk conductivity, $\sigma_b = \sigma_s/h_{\mathrm{eff}}$, where h_{eff} is the effective thickness of the 2DM. This approach can oftentimes create confusion due to the ambiguity contained in the definition of the thickness of an atomic monolayer. Moreover, the electromagnetic properties of 2DM can alternatively be described by the electric permittivity, ϵ, which is related to the optical conductivity σ_b via the following relation:

$$\epsilon(\omega) = \epsilon_0\left(1 + \frac{i\sigma_b}{\epsilon_0\omega}\right) = \epsilon_0\left(1 + \frac{i\sigma_s}{\epsilon_0\omega h_{\mathrm{eff}}}\right). \tag{6.2}$$

In the case of TMDC monolayer materials, their relative electric permittivity, $\epsilon_r(\omega) = \epsilon(\omega)/\epsilon_0$, can be described as a superposition of N Lorentzian functions:

$$\epsilon_r(\omega) = 1 + \sum_{k=1}^{N}\frac{f_k}{\omega_k^2 - \omega^2 - i\omega\gamma_k} = 1 + \sum_{k=1}^{N}\frac{f_k^E}{E_k^2 - E^2 - iE\gamma_k^E} \tag{6.3}$$

	WS_2			WSe_2			MoS_2			$MoSe_2$		
h_{eff}	6.18 Å			6.49 Å			6.15 Å			6.46 Å		
k	E_k(eV)	f_k^E(eV^2)	γ_k^E(eV)	E_k	f_k^E	γ_k^E	E_k	f_k^E	γ_k^E	E_k	f_k^E	γ_k^E
1	2.009	1.928	0.032	1.654	0.557	0.036	1.866	0.752	0.045	1.548	0.648	0.043
2	2.204	0.197	0.250	2.426	5.683	0.243	2.005	1.883	0.097	1.751	1.302	0.097
3	2.198	0.176	0.161	2.062	1.036	0.115	2.862	36.89	0.383	2.151	4.621	0.537
4	2.407	0.142	0.112	2.887	16.11	0.344	2.275	10.00	1.000	2.609	37.40	0.582
5	2.400	2.980	0.167	2.200	1.500	0.300	3.745	100.0	0.533	3.959	121.4	0.896
6	2.595	0.540	0.213	2.600	1.500	0.300		-			-	
7	2.644	0.050	0.171	3.800	70.00	0.700		-			-	
8	2.831	12.60	0.266	5.000	80.00	0.700		-			-	
9	3.056	8.765	0.240		-			-			-	
10	3.577	29.99	1.196		-			-			-	
11	5.078	49.99	1.900		-			-			-	
12	5.594	79.99	2.510		-			-			-	

Table 6.1: Model parameters for the relative permittivity of four TMDC monolayer materials in the multi-Lorentzian dispersion relation defined by Eq. (6.3). The oscillator strength, $f_k^E = \hbar^2 f_k$, the spectral resonance energy, $E_k = \hbar\omega_k$, and the spectral width, $\gamma_k^E = \hbar\gamma_k$, of the k-th oscillator are given in eV^2, eV, and eV,respectively.

where f_k, ω_k, and γ_k are the oscillator strength, the resonance frequency, and the spectral width of the k-th oscillator, respectively. The values of the model parameters for the four considered TMDC monolayers were determined by fitting Eq. (6.3) to the experimental data provided in Ref. [16] and are presented in Table 6.1 in terms of $f_k^E = \hbar^2 f_k$, $E_k = \hbar\omega_k$, and $\gamma_k^E = \hbar\gamma_k$. Note that the permittivity given by Eq. (6.3) fulfils the Kramers-Kronig relation; however, the numerical fitting is less accurate for $E = \hbar\omega > 3.1\,\text{eV}$, i.e. $\lambda < 0.4\,\mu\text{m}$, due to the limited range of the available experimental data.

The spectra of $\text{Re}\,(\sigma_s(\omega))$, a physical quantity related to the optical absorption in the material, and $\text{Im}\,(\sigma_s(\omega))$ are depicted in Fig. 6.3(a) and 6.3(b), receptively. For each TMDC, $\text{Re}\,(\sigma_s(\omega))$ exhibits spectral peaks at wavelengths specific to the particular material: the absorption peaks of $\text{Re}\,(\sigma_s(\omega))$ with highest and second highest wavelength for each material (shown for $MoSe_2$ as A and B, respectively) correspond to low-energy interband transitions at the $K(K')$-point of the first Brillouin zone due to splitting of the valence-band by spin-orbit coupling, whereas peaks at lower wavelengths correspond to

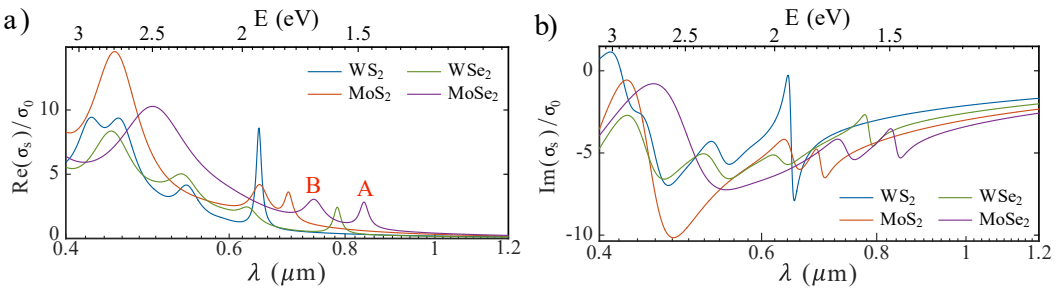

Figure 6.3: Frequency dependence of the real (a) and imaginary (b) parts of the sheet conductance $\sigma_s(\omega)$ of several TMDC monolayer materials, respectively.

higher energy interband transitions [16].

6.3.2 Nonlinear optical properties of 2D materials

The lattices of graphene and TMDC monolayers belong to different space symmetry groups, which means that each class of 2DM requires a separate treatment. The graphene lattice belongs to the $\mathscr{D}_{6h}$ point group, as illustrated in Fig. 6.4, which means that graphene is a centrosymmetric material and thus SHG is a forbidden nonlinear optical process. On the other hand, THG is an allowed, particularly strong process in graphene [14, 18], which makes it a suitable material for nonlinear optical applications. By contrast, TMDC monolayers belong to the $\mathscr{D}_{3h}$ point group [21] so that in this case SHG is the lowest-order nonlinear optical process.

The nonlinear optical response of graphene can be described by using the nonlinear optical conductivity tensor, $\boldsymbol{\sigma}_s^{(3)}$, which relates the nonlinear surface current density, $\mathbf{j}^{\mathrm{nl}}$, and the electric field, $\mathbf{E}$, at the fundamental frequency. Thus, if the graphene sheet is assumed to lie in the (x, y)-plane at z=0, the nonlinear current density, $\mathbf{J}^{\mathrm{nl}}$, can be written as:

$$\mathbf{J}^{\mathrm{nl}}(\mathbf{r}) = \mathbf{j}^{\mathrm{nl}}(\mathbf{r}_t)\delta(z), \tag{6.4}$$

where $\mathbf{r}_t$ is the position vector lying in the graphene plane. Then, the nonlinear surface current density can be written as:

$$\mathbf{j}_{\alpha}^{\mathrm{nl}}(\mathbf{r}_t) = \sigma_{s,\alpha}^{(3)} E_{\alpha}(\mathbf{r}_t)|E_t(\mathbf{r}_t)|^2, \tag{6.5}$$

where $\mathbf{E}_t$ is the electric field component lying in the plane of graphene. In this descrip-

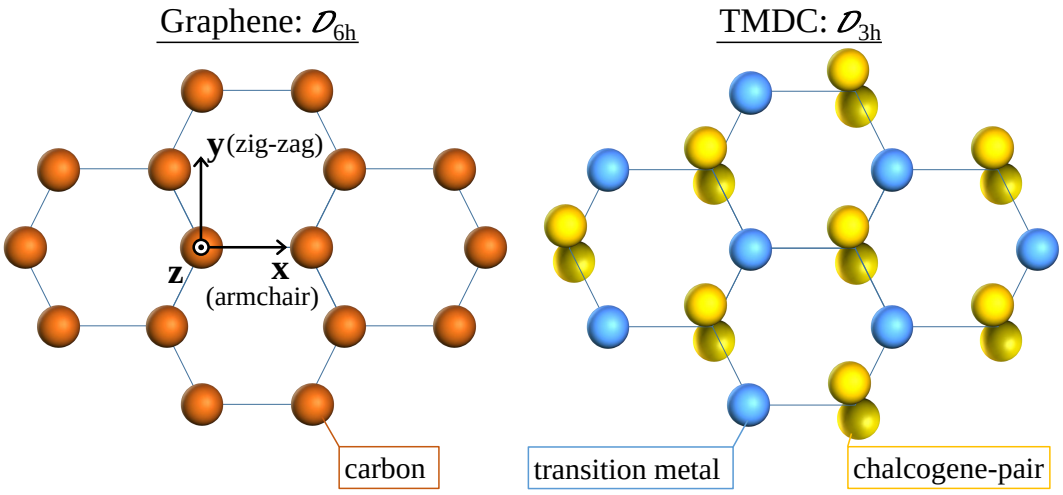

Figure 6.4: Schematic of the molecular structures of graphene (point group $\mathscr{D}_{6h}$, left) and the TMDC monolayer (point group $\mathscr{D}_{3h}$, right). The hexagonal lattice of graphene is centrosymmetric and thus SHG is forbidden. By contrast, the lattice of TMDC monolayers is non-centrosymmetric and as such SHG is allowed.

tion [14], the nonlinear current density lies in the plane of graphene and only depends on the tangential field components. As a result, $\sigma^{(3)}_{s,x} = \sigma^{(3)}_{s,y} = \sigma^{(3)}_s$ and $\sigma^{(3)}_{s,z} = 0$. A formula for $\sigma^{(3)}_s$, derived under the assumptions that electron-electron and electron-phonon scattering as well as thermal effects can be neglected, has been recently derived in Ref. [14] and reads:

$$\sigma^{(3)}_s(\omega) = \frac{i\sigma_0(\hbar v_F e)^2}{48\pi(\hbar\omega)^4} T\left(\frac{\hbar\omega}{2\varepsilon_F}\right), \tag{6.6}$$

In this equation, derived using perturbation theory, $T(x) = 17G(x) - 64G(2x) + 45G(3x)$,

$$G(x) = \ln\left|\frac{1+x}{1-x}\right| + i\pi\theta(|x|-1),$$

and $v_F = 3a_0\gamma_0/(2\hbar) \approx c/300$ is the Fermi velocity, with $a_0 = 1.42\,\text{Å}$ being the nearest-neighbour distance between carbon atoms in graphene and $\gamma_0 = 2.7\,\text{eV}$ the nearest-neighbour coupling constant. Figure 6.5(a) depicts $\sigma^{(3)}_s(\omega)$ and additionally highlights the lowest spectral peak at $\lambda = 1.033\,\mu\text{m}$ in the inset of the figure.

Similarly to the linear case, one can introduce a "bulk" nonlinear conductivity, $\boldsymbol{\sigma}^{(3)}_b = \boldsymbol{\sigma}^{(3)}_s/h_{\text{eff}}$. This nonlinear conductivity is particularly useful in experimental investigations of nonlinear optics of graphene because it is related to an effective bulk third-order nonlinear susceptibility, $\boldsymbol{\chi}^{(3)}_b$, the physical quantity that is usually measured experimentally. Using the fact that for harmonic fields $\mathbf{J}^{\text{nl}}(\mathbf{r}) = -i\omega\mathbf{P}^{\text{nl}}(\mathbf{r})$, where $\mathbf{P}^{\text{nl}}(\mathbf{r})$ is the complex coefficient of nonlinear polarisation, one can easily see that

$$\boldsymbol{\chi}^{(3)}_b = \frac{i}{\epsilon_0\Omega_t}\boldsymbol{\sigma}^{(3)}_b = \frac{i}{\epsilon_0\Omega_t h_{\text{eff}}}\boldsymbol{\sigma}^{(3)}_s, \tag{6.7}$$

where $\Omega_t = 3\omega_1$ is the frequency at the third-harmonic, with ω_1 being the fundamental frequency.

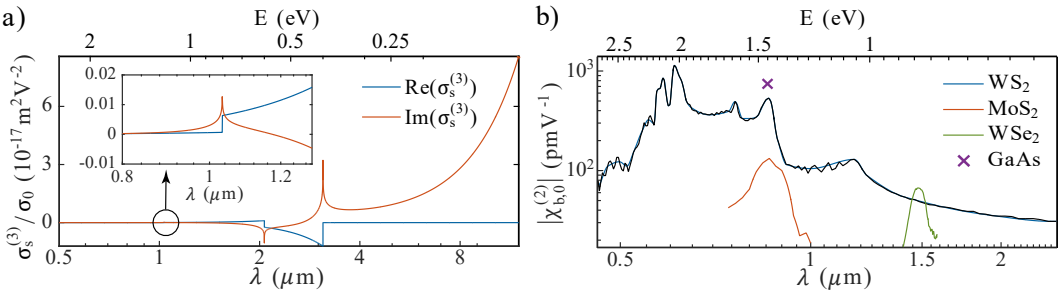

Figure 6.5: a) Real and imaginary part of the third order surface conductivity of graphene, $\sigma^{(3)}_s$. b) Spectral dependence of the effective bulk quadratic nonlinear susceptibility $|\chi^{(2)}_b|$ of three TMDC monolayer materials and the value of the dominant component of $\chi^{(2)}_{b,xyz}$ for GaAs, given for reference.

In contrast to graphene, TMDC monolayers are non-centrosymmetric (see Fig. 6.4) and therefore SHG is allowed [22, 23]. Based on the symmetry properties of their point group, $\mathscr{D}_{3h}$, it can be shown that the structure of their quadratically nonlinear susceptibility tensor, $\boldsymbol{\chi}_b^{(2)}$, yields only one independent, non-vanishing component [15, 22, 24].

$$\chi_{b,0}^{(2)} := \chi_{b,xxx}^{(2)} = -\chi_{b,xyy}^{(2)} = -\chi_{b,yxy}^{(2)} = -\chi_{b,yyx}^{(2)}, \tag{6.8}$$

where x is the armchair direction of the monolayer and y the orthogonal zig-zag-direction. The nonlinear surface conductivity tensor, $\boldsymbol{\sigma}^{(2)}$, has the same structure and is connected with the nonlinear susceptibility via a relation similar to Eq. (6.7):

$$\boldsymbol{\sigma}_s^{(2)} = -i\epsilon_0 \Omega_s h_{\text{eff}} \boldsymbol{\chi}_b^{(2)}, \tag{6.9}$$

where $\Omega_s = 2\omega_1$ is the second-harmonic frequency.

Numerical values for the effective thickness of the four TMDC monolayer materials are taken from Ref. [16] and are presented in Table 6.1. Although the tensorial structure expressed in Eq. (6.8) already qualitatively determines the nonlinear surface current,

$$j_\alpha^{\text{nl}}(\mathbf{r}_t) = \sum_{\beta,\gamma=x,y,z} \boldsymbol{\sigma}_{s,\alpha\beta\gamma}^{(2)} E_\beta(\mathbf{r}_t) E_\gamma(\mathbf{r}_t), \tag{6.10}$$

the numerical value of $\chi_{b,0}^{(2)}$ (and hence that of $\boldsymbol{\sigma}_s^{(2)}$) is of practical importance. Reliable values over a small spectral range exist for WS_2 [15], MoS_2 [24, 25], and WSe_2 [26] and are depicted in Fig. 6.5(b). Despite these materials consisting of just a single atomic layer, the largest values of $\chi_b^{(2)} = 1140\,\text{pm}\,\text{V}^{-1}$, $\chi_b^{(2)} = 132\,\text{pm}\,\text{V}^{-1}$, and $\chi_b^{(2)} = 67\,\text{pm}\,\text{V}^{-1}$, respectively, have similar magnitude as that of GaAs, $\chi_{b,xyz}^{(2)} = 740\,\text{pm}\,\text{V}^{-1}$, a medium with strong bulk quadratic susceptibility [23, 27].

It is also instructive to compare the strength of the SHG in TMDC monolayer materials to that of noble metals, as in both cases the SH is generated largely in a single atomic layer. To this end, a surface quadratic nonlinear susceptibility tensor, $\boldsymbol{\chi}_s^{(2)} = h_{\text{eff}} \boldsymbol{\chi}_b^{(2)}$ is introduced. This is then compared to the surface quadratic nonlinear susceptibility tensor of Au and Ag. The highest value of $\boldsymbol{\chi}_s^{(2)}$ of WS_2 is $\chi_s^{(2)} = 7 \times 10^{-19}\,\text{m}^2\,\text{V}^{-1}$ and is hence comparable to the dominant tensor component, $\chi_{s,\perp\perp\perp}^{(2)}$, for Ag and Au, i.e. $\chi_{s,\perp\perp\perp}^{(2)} = 1.59 \times 10^{-18}\,\text{m}^2\,\text{V}^{-1}$ and $\chi_{s,\perp\perp\perp}^{(2)} = 1.35 \times 10^{-18}\,\text{m}^2\,\text{V}^{-1}$ [28], respectively.

6.4 Mathematical description of frequency generation in periodically patterned 2D materials

This section will extend the RCWA to the case of nonlinear optical interaction of light with periodic structures consisting of 2DMs embedded in dielectric or metallic patterned media.

Nonlinear optical effects in the bulk-part of periodic structures are not considered here but will be treated in Chapter 7. Moreover, there are numerical formulations of the RCWA to accurately calculate these bulk-nonlinear optical effects that are compatible [29–31] to the proposed method here.

To this end, the undepleted pump approximation will be used as a means to introduce the nonlinear optical interactions in the governing Maxwell equations. Then, a modified boundary condition for interfaces incorporating conductive, potentially nonlinear 2DMs, is considered and a numerically stable $\mathcal{S}$-matrix algorithm for propagation of nonlinear fields generated by the monolayers of 2DMs in multilayered periodic structures is derived to complete the proposed algorithm. Importantly, the thickness of the 2DMs does not enter in our numerical method, thus removing an ambiguity present in other numerical methods currently used to describe these materials [4–6].

6.4.1 Optical harmonic generation in the undepleted pump approximation

This section introduces the physical and mathematical model for multi-frequency, nonlinear optical interaction in the so-called undepleted pump approximation. For a more complete description of this approach we refer the reader to Ref. [23]. To this end, assume that the real electric field, $\boldsymbol{\mathscr{E}}(\mathbf{r},t)$, as a function of position, $\mathbf{r}$, and time, t, is composed of contributions of N_F+1 monochromatic components with pairwise different optical frequencies, ω_n, $n=0,\ldots,N_F$, i.e.

$$\boldsymbol{\mathscr{E}}(\mathbf{r},t)=\frac{1}{2}\sum_{n=0}^{N_F}\mathbf{E}^{(\omega_n)}(\mathbf{r})\exp(-i\omega_n t)+c.c., \tag{6.11}$$

where $\mathbf{E}^{(\omega_n)}(\mathbf{r})$ is the complex amplitude of the wave with frequency, ω_n, and $c.c.$ denotes the complex conjugate of the preceding expression. The fields with frequencies ω_n, $n=1,\ldots,N_F$, are assumed to be excitation (pump) fields, whereas the field at $\Omega:=\omega_0$ is a higher-harmonic, nonlinearly generated field. This nonlinear optical field depends on the specific nonlinear optical process under investigation. Similar expressions are assumed for the other electromagnetic quantities. Particularly, the polarisation $\boldsymbol{\mathscr{P}}(\mathbf{r},t)$ is then expressed as:

$$\boldsymbol{\mathscr{P}}(\mathbf{r},t)=\frac{1}{2}\sum_{n=0}^{N_F}\mathbf{P}^{(\omega_n)}(\boldsymbol{\mathscr{E}};\mathbf{r})\exp(-i\omega_n t)+c.c., \tag{6.12}$$

where $\mathbf{P}^{(\omega_n)}(\boldsymbol{\mathscr{E}};\mathbf{r})$ encodes a – possibly nonlinear – functional relation between the total, time dependant field $\boldsymbol{\mathscr{E}}$ and the polarisation with ω_n-harmonic time dependence at position $\mathbf{r}$.

In most situations of practical interest, depending on particular physical conditions,

only certain nonlinear polarisations are generated with significant strength. Moreover, in the case of strong excitation fields or when the induced nonlinear polarisations are weak, the following assumption can be made:

$$\mathbf{P}^{(\omega_n)}(\boldsymbol{\mathscr{E}};\mathbf{r}) = \mathbf{P}^{(\omega_n)}(\mathbf{E}^{(\omega_n)};\mathbf{r}) = \epsilon_0\chi^{(1)}(\mathbf{r};\omega_n)\mathbf{E}^{(\omega_n)}(\mathbf{r}).$$

This means that the polarisations at the frequencies of the pump fields, ω_n, $n = 1,\ldots,N_F$, are solely determined by the corresponding linear optical susceptibility, $\chi^{(1)}(\omega_n)$, of the optical medium and the electric field at frequency ω_n. The polarisation at the nonlinear frequency, Ω, however, consists of both the linear polarisation at Ω, which is proportional to the nonlinear field, $\mathbf{E}^{(\Omega)}(\mathbf{r})$, and a nonlinear polarisation $\mathbf{P}^{(\mathrm{nl},\Omega)}(\mathbf{E}^{(\omega_{n\neq 0})};\mathbf{r})$, which can incorporate the electric field from all pump fields with frequencies ω_n, $n > 0$:

$$\mathbf{P}^{(\Omega)}(\boldsymbol{\mathscr{E}};\mathbf{r}) = \epsilon_0\chi^{(1)}(\Omega)\mathbf{E}^{(\Omega)}(\mathbf{r}) + \mathbf{P}^{(\mathrm{nl},\Omega)}(\mathbf{E}^{(\omega_{n\neq 0})};\mathbf{r}). \tag{6.13}$$

Under this assumption, the optical fields at the pump frequencies, $\omega_1,\ldots,\omega_{N_F}$, are not altered, or in a narrower sense, depleted by the nonlinear processes, hence this assumption is called the undepleted pump approximation. The particular form of the nonlinear polarisation, $\mathbf{P}^{(\mathrm{nl},\Omega)}(\mathbf{E}^{(\omega_{n\neq 0})};\mathbf{r})$ depends on the nonlinear process under consideration, such as sum- or difference-frequency generation, and second- or third-harmonic generation. Specific expressions for $\mathbf{P}^{(\mathrm{nl},\Omega)}(\mathbf{E}^{(\omega_{n\neq 0})};\mathbf{r})$ have been given in the previous section for 2DMs.

The algorithmic appeal of this approximation is the possibility to obtain a general, one-way coupled calculation scheme, which only requires the solution of N_F homogeneous linear optical problems and one affine linear problem with electrical sources. In particular, the numerical algorithm consists of the following steps: In the step, *i*) (*linear calculations*) one calculates the fields $\mathbf{E}^{(\omega_n)}$ for the pump frequencies, ω_n, $n = 1,\ldots,N_F$. In the second step, *ii*) (*polarisation evaluation*) one evaluates the nonlinear polarisation, $\mathbf{P}^{(\mathrm{nl},\Omega)}(\mathbf{E}^{(\omega_{n\neq 0})};\mathbf{r})$, for the particular nonlinear process under consideration. Finally, in the third step *iii*), (*nonlinear calculation*), one calculates the generated nonlinear electric field, $\mathbf{E}^{(\Omega)}$.

Given this generic three-step algorithm, the remaining part of this section will describe the numerical implementation of the computational steps *i*) and *iii*) for the particular case of higher-harmonic generation in multilayered periodic structures, which contain periodically patterned monolayers of 2DMs that can exhibit quadratic or cubic optical nonlinearity.

6.4.2 RCWA: modal expression for fields in bulk periodic regions

The main result that is necessary for the incorporation of nonlinear 2DMs into RCWA is that the solutions in each of the bulk layers are given by a modal expansion: For each frequency component $\omega = \omega_n$, $n = 0,\ldots,N_F$, the electric field $\mathbf{E}^{(\omega)}$ is given as a superposition

of modes, where each mode profile itself is given as a Fourier series. This was derived in Section 3.4 and the main result is repeated here, where the superscript (ω) is dropped, because the description of the modal form of the electromagnetic fields is independent of whether a pump or generated frequency is considered.

The electromagnetic fields in each layer, denoted with superscript "L", are expressed in a linear combination of upward- and downward-propagating modes with coefficients $c_m^{(L,\pm)}$, where superscripts "$+$" and "$-$" refer to upward and downward propagation, respectively. Each mode is fully described by its complex propagation constant, $\nu_m^{(L,\pm)}$, which determines the z-dependence, and the z-independent Fourier series coefficients $E_{\alpha,n}^{(L,m,\pm)}$ and $H_{\alpha,n}^{(L,m,\pm)}$, for $\alpha = x, y, z$. These coefficients determine the transverse profile of mode m via the Fourier series reconstruction operator, $\mathscr{R}(\cdot)$, from Eq. (3.20).

This Fourier series reconstruction of the total electric and magnetic fields are hence expressed by:

$$E_\alpha^{(L)}(x,y,z) = \mathscr{R}\left([E_\alpha^{(L)}(z)]\right)(x,y), \tag{6.14a}$$

$$H_\alpha^{(L)}(x,y,z) = \mathscr{R}\left([H_\alpha^{(L)}(z)]\right)(x,y), \tag{6.14b}$$

with

$$\begin{pmatrix} [E_\alpha^{(L)}(z)] \\ [H_\alpha^{(L)}(z)] \end{pmatrix} = \begin{bmatrix} \mathsf{E}_\alpha^{(L,+)} & \mathsf{E}_\alpha^{(L,-)} \\ \mathsf{H}_\alpha^{(L,+)} & \mathsf{H}_\alpha^{(L,-)} \end{bmatrix} \begin{bmatrix} \mathsf{V}^{(L,+)}(z) & 0 \\ 0 & \mathsf{V}^{(L,-)}(z) \end{bmatrix} \begin{pmatrix} \mathbf{c}^{(L,+)} \\ \mathbf{c}^{(L,-)} \end{pmatrix}. \tag{6.15}$$

Here, the m-th column of $\mathsf{E}_\alpha^{(L,\pm)}$ and $\mathsf{H}_\alpha^{(L,\pm)}$ is the vector of Fourier coefficients $[E_\alpha^{(L,m,\pm)}]$ and $[H_\alpha^{(L,m,\pm)}]$, respectively, of the α-component of the m-th mode. The propagation matrix $\mathsf{V}^{(L,\pm)}(z)$ is diagonal with entries $\mathsf{V}_{mm}^{(L,\pm)}(z) = \exp\left(ik_0\nu_m^{(L,\pm)}(z - z^{(L,\mp)})\right)$, where $z^{(L,\pm)}$ is the bottom/top z-coordinate of layer L. This means that the electromagnetic fields in each layer are fully determined by their coefficients $\mathbf{c}^{(L,\pm)}$.

6.4.3 Modelling 2D materials via RCWA by means of boundary conditions

The optical response of 2DMs can chiefly be described in two different ways. One direct approach in RCWA and other numerical methods is to model the monolayer material as a periodic, very thin film, similar to the layers in which the periodic bulk components are decomposed. To this end, a physical effective thickness h_{eff} of the 2DM under consideration has to be chosen as well as its permittivity. This is a rather questionable and inefficient approach, for three main reasons. First, the choice of h_{eff} is somewhat ambiguous, both because the thickness of an atomic monolayer does not have a clear meaning in the classical physics context and due to the fact that the experimentally measured value of this thickness varies considerably for all 2DMs. Second, this approach suffers from numerical artefacts and slow convergence, as reported in recent works [8, 11]. This is understandable as the

thickness of the monolayer is much smaller than the optical wavelength, $h_{\text{eff}} \ll \lambda$. Finally, it is computationally costly to find RCWA modes of the 2DMs since this requires the numerical solution of an eigenvalue problem, see Section 3.4.

An alternative procedure for describing sheets of structured 2DMs is introduced in this chapter and consists of modelling such 2DM-sheets by means of a surface conductance, which enters the algorithm via the electromagnetic boundary conditions between adjacent bulk layers. Thus, consider the schematic of the multilayer structure in Fig. 6.6(a). At a horizontal plane with $z = z_s$, which is located between two adjacent bulk layers, the following relations of the tangential $\mathbf{E}$ and $\mathbf{H}$ fields in the top region (superscript A) and the bottom region (superscript B) have to be fulfilled [32]:

$$\hat{\mathbf{n}} \times \left(\mathbf{E}^{(\omega,A)}(x,y,z_s) - \mathbf{E}^{(\omega,B)}(x,y,z_s) \right) = 0, \tag{6.16a}$$

$$\hat{\mathbf{n}} \times \left(\mathbf{H}^{(\omega,A)}(x,y,z_s) - \mathbf{H}^{(\omega,B)}(x,y,z_s) \right) = \mathbf{j}^{(\omega)}(x,y), \tag{6.16b}$$

where $\hat{\mathbf{n}} = (0,0,1)$ denotes the unit vector normal to the horizontal plane, pointing towards region A. The first of these equations expresses the continuity of the tangential components of $\mathbf{E}$ at the interface. The second equation warrants further discussion. Thus, in the presence of a 2DM at $z = z_s$, the tangential component of $\mathbf{H}$ is discontinuous, its variation across the i-th interface being given by the surface current, $\mathbf{j}^{(\omega)}(x,y) = \mathbf{j}^{(\omega,\text{lin})}(x,y) + \mathbf{j}^{(\omega,\text{nl})}(x,y)$.

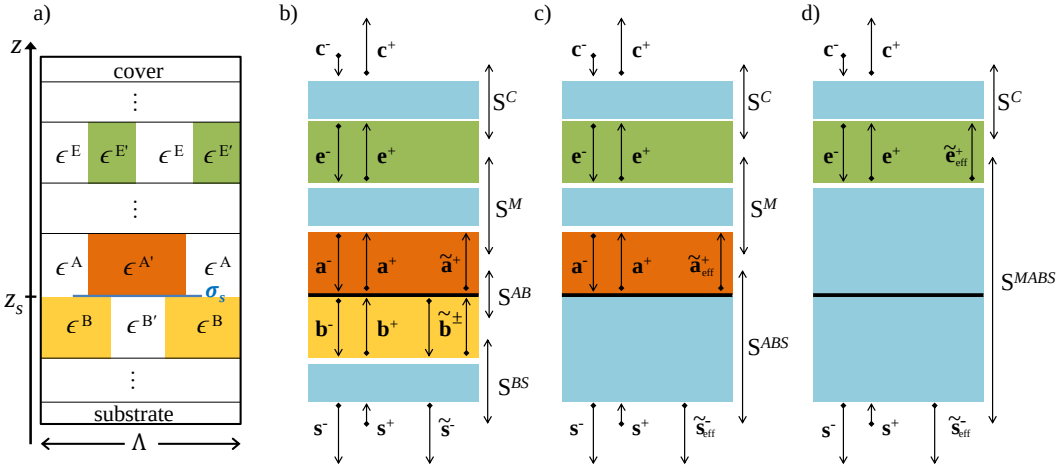

Figure 6.6: a) Schematic of a multilayer structure located between cover and substrate containing a 2DM sheet with surface conductance σ_s, the neighbouring layers A and B, and the evaluation layer, E, in which the electromagnetic field is to be determined. b) Linear ($\mathbf{c}^{\pm}, \mathbf{e}^{\pm}, \ldots$) and nonlinear ($\tilde{\mathbf{a}}^{+}, \tilde{\mathbf{b}}^{\pm}, \tilde{\mathbf{s}}^{-}$) mode coefficients in the respective layers and the $\mathcal{S}$-matrices that connect them. c) Combination of layers B and S yields the combined $\mathcal{S}$-matrix S^{ABS} and effective nonlinear coefficients $\tilde{\mathbf{a}}^{+}_{\text{eff}}$ and $\tilde{\mathbf{s}}^{-}_{\text{eff}}$. d) Repeated combination of $\mathcal{S}$-matrices allows the calculation of the field coefficients, $\mathbf{e}$, from the effective nonlinear mode coefficients $\tilde{\mathbf{e}}^{+}_{Eff}$.

In particular, the total surface current contains a linear component given by,

$$\mathbf{j}^{(\omega,\mathrm{lin})}(x,y) = \sigma_s^{(\omega)}(x,y)\mathbf{E}^{(\omega,s)}(x,y), \tag{6.17}$$

and a nonlinear surface current, $\mathbf{j}^{(\omega,\mathrm{nl})}(x,y)$. The linear surface current depends only on the electric field at the interface, $\mathbf{E}^{(\omega,s)}(x,y,z_s)$, at the same frequency, ω, and the distribution of the sheet conductivity at the interface. The nonlinear surface current, on the other hand, is assumed to be non-vanishing only at the frequency $\omega = \omega_0$ and generally depends on the electric field at all the other frequencies, $\omega = \omega_n$, $n = 1\ldots, N_F$.

One can easily see that $\mathbf{j}^{(\omega)}(x,y)$ is a pseudo-periodic function of the x- and y-coordinate, hence it is determined by its Fourier coefficient vector, $[\mathbf{j}^{(\omega)}]$. Special attention is, however, necessary when one calculates the Fourier coefficients of the linear current, $[\mathbf{j}^{(\omega,\mathrm{lin})}]$, as in the real space it is given by a product of two periodic functions, $\sigma_s^{(\omega)}(x,y)$ and $\mathbf{E}^{(\omega,s)}(x,y)$. This issue is another occurrence of the Fourier factorisation problem Section 3.3.3 and it must be addressed appropriately to obtain high accuracy and fast convergence of the proposed method here as it relies on Fourier series representation of the EM solution [33–36].

To find the correct factorisation rule for Eq. (6.17), the continuity properties of the two factors in the r.h.s. of this equation as well as that of the product have to be investigated. To this end, consider Fig. 6.7, which depicts the $z = z_s$ plane that comprises a surface conductivity, $\sigma_s^{(\omega)}(x,y) = \chi_{2\mathrm{DM}}(x,y)\sigma_s(\omega)$. Hereby, $\chi_{2\mathrm{DM}}(x,y)$ denotes the characteristic function of the 2DM distribution and $\sigma_s(\omega)$ is the sheet conductance of the 2DM, e.g. Eq. (6.1) for graphene. At each point, $\tilde{\mathbf{r}}$, of the 1D boundary Γ of the 2DM, a local coordinate system is introduced and defined by the orthonormal vectors $\mathbf{t}_n$, $\mathbf{t}_t$, and $\mathbf{e}_z$, where $\mathbf{t}_n$ and $\mathbf{t}_t$ are the in-plane unit vectors normal and tangential to the 1D contour, Γ, respectively, at the point $\tilde{\mathbf{r}}$ and $\mathbf{e}_z$ is the unit vector along the z-axis. One can assume that the sheet conductance function is smooth along the $\mathbf{t}_t$ direction, yet it is discontinuous along the $\mathbf{t}_n$

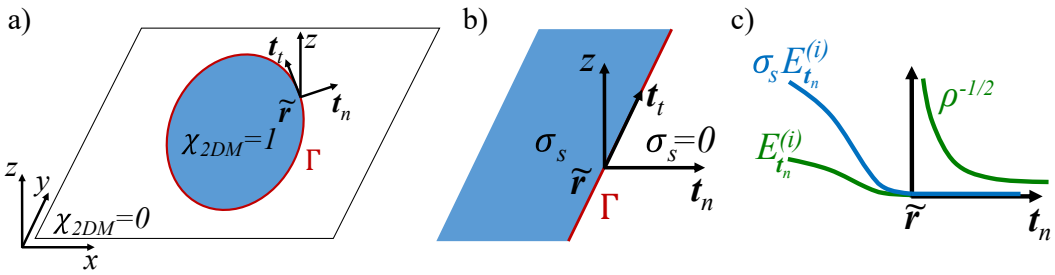

Figure 6.7: a) Cross-section of the periodic structure through the plane $z = z_s$ containing the 2DM distribution $\chi_{2\mathrm{DM}}(x,y)$ with the boundary contour, Γ, showing a unit cell and the local coordinate system at a generic location, $\tilde{\mathbf{r}}$. b) Local coordinate system at $\tilde{\mathbf{r}}$. c) Qualitative behaviour of the surface quantities $E_{\mathbf{t}_n}^{(i)}$ and $j_{\mathbf{t}_n}^{(\omega,\mathrm{lin})} = \sigma_s^{(\omega)} E_{\mathbf{t}_n}^{(\omega,s)}$

direction, as per Fig. 6.7(b). The continuity relation, Eq. (6.16a), of the tangential electric field at $z = z_s$ allows one to define a tangential surface field, $E^{(\omega,s)}_{\mathbf{t}_n}(x, y)$, as the limit of the volumetric electric fields, $E^{(\omega)}_{\mathbf{t}_n}(x, y, z)$, from either side of the $z = z_s$ plane:

$$E^{(\omega,s)}_{\mathbf{t}_n}(x, y) := \lim_{z\to z_s} E^{(\omega)}_{\mathbf{t}_n}(x, y, z). \tag{6.18}$$

The electromagnetic near-field in the vicinity of such a conductive sheet can be calculated analytically [37] and the main results relevant to this work is summarised here: *i*) The $\mathbf{t}_n$-component of the surface current vanishes at $\tilde{\mathbf{r}}$ as

$$\lim_{\rho\to 0^+} j^{(\omega,\mathrm{lin})}_{\mathbf{t}_n}(\tilde{\mathbf{r}} - \rho\mathbf{t}_n) = \lim_{\rho\to 0^+} C\sqrt{k\rho}e^{ik\rho} = 0. \tag{6.19}$$

Since $j^{(\omega,\mathrm{lin})}_{\mathbf{t}_n}(\tilde{\mathbf{r}} + \rho\mathbf{t}_n) = 0$ for $\rho > 0$, $j^{(\omega,\mathrm{lin})}_{\mathbf{t}_n}(\tilde{\mathbf{r}})$ is continuous around $\tilde{\mathbf{r}}$. *ii*) The $\mathbf{t}_n$-component of the surface field, $E^{(\omega,s)}_{\mathbf{t}_n}(\tilde{\mathbf{r}})$, is discontinuous at $\tilde{\mathbf{r}}$. More specifically,

$$\lim_{\rho\to 0^+} E^{(\omega,s)}_{\mathbf{t}_n}(\tilde{\mathbf{r}} - \rho\mathbf{t}_n) = 0, \tag{6.20}$$

that is $E^{(\omega,s)}_{\mathbf{t}_n}(\tilde{\mathbf{r}})$ inside the 2DM vanishes near the boundary, whereas it diverges when approaching the interface from outside the 2DM,

$$\lim_{\rho\to 0^+} E^{(\omega,s)}_{\mathbf{t}_n}(\tilde{\mathbf{r}} + \rho\mathbf{t}_n) = \lim_{\rho\to 0^+} C(k\rho)^{-1/2} = \infty. \tag{6.21}$$

Finally, *iii*) the in-plane tangential component $E^{(\omega,s)}_{\mathbf{t}_t}(\tilde{\mathbf{r}})$ of the interfacial electric field is continuous and bounded at $\tilde{\mathbf{r}}$. Findings *i*) and *ii*) are schematically depicted in Fig. 6.7(c).

These results imply that $j^{(\omega,\mathrm{lin})}_{\mathbf{t}_n}(x, y)$ is a continuous function but both factors, $\sigma^{(\omega)}_s(x, y)$ and $E^{(\omega,s)}_{\mathbf{t}_n}(x, y)$, in Eq. (6.17) are discontinuous. Therefore, Laurent's product rule (Theorem *I* in Section 3.3.3) cannot be applied, as it would result in spurious field oscillations near the contour Γ and slow convergence of the algorithm. The inverse rule (Theorem *III* in Section 3.3.3) is also not applicable, as the first factor, $\sigma^{(\omega)}_s(x, y)$, vanishes at some regions of the unit cell, outside of the areas occupied by the 2DM. To overcome this problem, a small but non-zero surface conductivity, $\sigma_{s,\mathrm{add}} \neq 0$, is added at the interface regions where there is no 2DM with non-vanishing conductivity. As a result, this modified sheet conductance distribution in the unit cell is defined as:

$$\tilde{\sigma}^{(\omega)}_s(x, y) := \chi_{2\mathrm{DM}}(x, y)\sigma_s(\omega) + [1 - \chi_{2\mathrm{DM}}(x, y)]\,\sigma_{s,\mathrm{add}}. \tag{6.22}$$

As we have seen in Section 6.3.1, $\sigma_s(\omega)$ is a function of frequency and can vary by orders of magnitude. Therefore, it is natural to scale the additional conductivity, $\sigma_{s,\mathrm{add}}$,

relative to the absolute value of the physical conductivity of the 2DM:

$$\sigma_{s,\mathrm{add}}(\omega,\eta) = -i\eta|\sigma_s(\omega)|, \tag{6.23}$$

where η is a small proportionality constant. This choice ensures in addition that $\sigma_{s,\mathrm{add}}$ is a negative imaginary quantity, so that one does not introduce artificial losses in the system but only a vanishingly small phase change at the interface. However small this effect is, it introduces an error and hence its quantitative contribution to the optical response of the photonic structure must be investigated. The influence of $\sigma_{s,\mathrm{add}}$, and implicitly that of η, on the calculated near- and far-fields by the proposed method will be investigated in Section 6.5.

An alternative justification of the nonzero conductivity model presented here can be provided by the application of Ampere's law for a loop around the interface, an approach that has been recently introduced elsewhere [8, 11]. This alternative model also amounts to using a modified conductivity at the interface similar to that in Eq. (6.22).

In order to complete the derivation of the correct Fourier factorisation of the surface current defined by Eq. (6.17) for $\tilde{\sigma}_s^{(\omega)}(x,y) \neq 0$, let $\mathbf{N}(x,y)$ denote a normal vector field, namely a vector field that is normal to the 1D contour Γ at any point of the interface and is continuously extended into the regions away from Γ. This can be done analytically for certain cross sections and ways to automatically generate a NVF for arbitrary cross-sections have been discussed already in Section 5.3.2.

This allows one to express the factorisation of $[\mathbf{j}^{(\omega,\mathrm{lin})}]$ in terms of the conductivity distribution coefficients, $\left[\!\left[\tilde{\sigma}_s^{(\omega)}\right]\!\right]$ and $\left[\!\left[1/\tilde{\sigma}_s^{(\omega)}\right]\!\right]$, and tangential field coefficients, $[\mathbf{E}^{(\omega,s)}]$, in such a way that the normal $\mathbf{t}_n$-component of $\mathbf{j}^{(\omega,\mathrm{lin})}(x,y)$ is decomposed using the inverse rule, and its tangential $\mathbf{t}_t$-component is decomposed using the regular product rule:

$$[j_\alpha^{(\mathrm{lin})}] = \sum_{\beta=x,y} \Delta\mathsf{N}_{\alpha,\beta}[E_\beta^{(\omega,s)}], \tag{6.24}$$

where the conductivity difference matrix is given by

$$\begin{aligned}\Delta\mathsf{N}_{\alpha\beta} = \delta_{\alpha\beta}\left[\!\left[\tilde{\sigma}^{(\omega)}\right]\!\right] &+ \frac{1}{2}[\![N_\alpha N_\beta]\!]\left(\left[\!\left[1/\tilde{\sigma}_s^{(\omega)}\right]\!\right]^{-1} - \left[\!\left[\tilde{\sigma}_s^{(\omega)}\right]\!\right]\right) \\ &+ \frac{1}{2}\left(\left[\!\left[1/\tilde{\sigma}_s^{(\omega)}\right]\!\right]^{-1} - \left[\!\left[\tilde{\sigma}_s^{(\omega)}\right]\!\right]\right)[\![N_\alpha N_\beta]\!].\end{aligned} \tag{6.25}$$

This procedure is analogous to the normal vector field factorisation approach for the displacement field in the regular RCWA for bulk materials given in Section 3.3.5.3. It should be noted that the fast Fourier factorisation by *zig-zag* approximation, Section 3.3.5.1, for linear interaction of light with 2D materials has been derived in Ref. [11].

The calculation of $\mathbf{j}_{\mathbf{t}_n}^{(\omega_0,\mathrm{nl})}(x,y)$ as a function of $\mathbf{E}$ requires the electric near-field distribution at the interface, which is difficult to calculate accurately in the RCWA, even when the correct Fourier factorisation rules are employed, as we have seen in Chapter 4 and was noted earlier in Refs. [38–40]. The main insight into the accuracy of near-field calculations is that RCWA relies on the accurate description of continuous quantities, as they can be readily expanded in Fourier series. This can be exploited to achieve an accurate interface-field description: knowing that the $\mathbf{t}_t$-component of the interface electric field, $\mathbf{E}_{\mathbf{t}_t}^{(\omega,s)}$, is continuous, it is convenient to evaluate this component directly by Fourier series reconstruction. Since, however, its $\mathbf{t}_n$-component is discontinuous (and in fact singular), it can only be poorly represented by its Fourier series, so that the reconstructed field $\mathbf{E}^{(\omega,s)}$ will experience unphysical oscillations and the Gibbs phenomenon.

A more well-behaved quantity is the surface current, which is continuous and hence easily described by its Fourier series. With the definition of the Fourier coefficients of the vectorial normal surface current and tangential surface fields as

$$[\mathbf{E}_{\mathbf{t}_t}^{(\omega,s)}] = [\![\mathsf{I} - \mathbf{N}\mathbf{N}^T]\!][\mathbf{E}_{\mathbf{t}}^{(s)}], \tag{6.26a}$$

$$[\mathbf{j}_{\mathbf{t}_n}^{(\omega)}] = \frac{1}{2}\left([\![\mathbf{N}\mathbf{N}^T]\!] \left[\!\left[1/\tilde{\sigma}_s^{(\omega)}\right]\!\right]^{-1} + \left[\!\left[1/\tilde{\sigma}_s^{(\omega)}\right]\!\right]^{-1} [\![\mathbf{N}\mathbf{N}^T]\!] \right) [\mathbf{E}_{\mathbf{t}}^{(\omega,s)}], \tag{6.26b}$$

one obtains the reconstructed field at the interface as:

$$\mathbf{E}_{\mathbf{t}}^{(\omega,s)}(x,y) = \mathscr{R}\left([\mathbf{E}_{\mathbf{t}_t}^{(\omega,s)}]\right)(x,y) + \mathscr{R}\left([\mathbf{j}_{\mathbf{t}_n}^{(\omega)}]\right)(x,y)/\sigma_s(\omega). \tag{6.27}$$

Here, all Fourier series represent functions continuous at the in-plane boundaries of the 2DM. The fact that this approach only allows a reconstruction of the near-field at the 2DM structure, where $\chi_{\mathrm{2DM}}\sigma_s \neq 0$, is not particularly concerning since the nonlinear surface current is a priori nonvanishing there only.

6.4.4 Inhomogeneous $\mathscr{S}$-matrix formalism

The last section revisited the modal expansion of the EM fields everywhere in and around the grating structure and cast them in a concise matrix form Eq. (6.15). It remains to determine the coefficients $\mathbf{c}^{\pm}$ in every computational layer to obtain the EM fields. This is achieved by fulfilling the EM boundary conditions Eq. (6.16b) between all computational layers and will be cast in a flexible inhomogeneous $\mathscr{S}$-matrix formalism, which is introduced now.

To this end, consider again the multilayer structure and the computational variables defining the corresponding electromagnetic field schematically illustrated in Figs. 6.6(a) and 6.6(b), respectively. The optical structure consists of exactly one 2DM-sheet at $z = z_s$, defined by its distribution of conductivity, and an arbitrary number of bulk layers, defined by their respective electrical permittivity distribution and identified by their superscript.

Three of the layers are of particular interest, namely the two computational layers directly enclosing the 2DM-sheet, identified by "A" and "B", and the evaluation layer, "E", which can be any computational bulk layer in the grating structure. Additionally, "C" and "S" identify the cover and substrate layers, respectively. These capital letters are used to identify the mode-shape matrices, $\mathsf{E}_\alpha^{(L,\pm)}$ and $\mathsf{H}_\alpha^{(L,\pm)}$, and the propagation matrix, $\mathsf{V}^{(L,\pm)}(z)$, in each layer, $L \in \{A, B, E, C, S\}$. The vector of mode coefficients in each layer, L, is denoted by the respective bold lower-case letter, $\mathbf{a}^\pm, \mathbf{b}^\pm, \mathbf{e}^\pm, \mathbf{c}^\pm$, and $\mathbf{s}^\pm$.

At each of the pump frequencies, plane wave incidence is assumed, i.e. the incoming cover and substrate coefficients, $\mathbf{c}^-$ and $\mathbf{s}^+$, respectively, are given and at least one of their entries is non-vanishing. At the generated frequency, no field is assumed to be incident, hence $\mathbf{s}^- = \mathbf{c}^+ = 0$. Moreover, the higher-harmonic optical field is generated due to the nonlinear surface current $\mathbf{j}^{(\omega,\mathrm{nl})}(x, y)$. With this set-up, the goal of the remaining derivation is to determine the mode coefficients, $\mathbf{e}^\pm$, in the evaluation layer, E. Since the layer E was chosen arbitrarily, this suffices to determine the mode coefficients, and hence the EM fields in any layer.

Since the fields in layers A and B are expressed in their Fourier Series according to Eq. (6.14), the boundary conditions Eq. (6.16) can be expressed in terms of the Fourier series coefficients of the fields and read:

$$[E_{\mathbf{t}}^{(A)}(z_s)] = [E_{\mathbf{t}}^{(B)}(z_s)], \tag{6.28a}$$

$$[H_{\mathbf{t}}^{(A)}(z_s)] - [H_{\mathbf{t}}^{(B)}(z_s)] = [\delta H_{\mathbf{t}}^{(s)}]. \tag{6.28b}$$

Hereby, $\delta H_{\mathbf{t}}^{(s)} = (j_y, -j_x)^T$ denotes the variation of the tangential components of $\mathbf{H}$ across the surface at $z = z_s$ and the superscript, ω, is dropped to simplify the notation.

The fields $[E_{\mathbf{t}}^{(A)}(z_s)]$ and $[H_{\mathbf{t}}^{(A)}(z_s)]$ can be expressed in terms of their mode coefficients $\mathbf{a}^\pm$ and $\mathbf{b}^\pm$ via Eq. (6.15). This same procedure is performed with the linear surface current according to the factorisation rule Eq. (6.24), which relies on the tangential field at $z = z_s$. Due to the continuity of the tangential surface field, in principle, the values from either side A or B, or their average could be taken, as they are all equal. The latter is chosen here, thus yielding $[E_{\mathbf{t}}^{(s)}] = ([E_{\mathbf{t}}^{(A)}(z_s)] + [E_{\mathbf{t}}^{(B)}(z_s)])/2$.

Using the factorisation rule given in Eq. (6.25), one obtains for the α-component of the linear surface current:

$$[j_\alpha^{(\mathrm{lin})}] = \frac{1}{2} \sum_{\beta=x,y} \Delta\mathsf{N}_{\alpha\beta} \times \left\{ \begin{bmatrix} \mathsf{E}_\beta^{(A,+)} & \mathsf{E}_\beta^{(B,-)} \end{bmatrix} \begin{bmatrix} \mathsf{V}^{(A,+)}(z_s) & 0 \\ 0 & \mathsf{V}^{(B,-)}(z_s) \end{bmatrix} \begin{pmatrix} \mathbf{a}^+ \\ \mathbf{b}^- \end{pmatrix} \right. \tag{6.29}$$

$$\left. + \begin{bmatrix} \mathsf{E}_\beta^{(A,-)} & \mathsf{E}_\beta^{(B,+)} \end{bmatrix} \begin{bmatrix} \mathsf{V}^{(A,-)}(z_s) & 0 \\ 0 & \mathsf{V}^{(B,+)}(z_s) \end{bmatrix} \begin{pmatrix} \mathbf{a}^- \\ \mathbf{b}^+ \end{pmatrix} \right\}$$

Defining the block matrix $\mathsf{G}_{\mathbf{t}}^{(L,\pm)}$ for layers A and B as

$$\mathsf{G}_{\mathbf{t}}^{(L,\pm)} = \frac{1}{2}\begin{bmatrix} +\Delta N_{yx}\mathsf{E}_x^{(L,\pm)} + \Delta N_{yy}\mathsf{E}_y^{(L,\pm)} \\ -\Delta N_{xx}\mathsf{E}_x^{(L,\pm)} - \Delta N_{xy}\mathsf{E}_y^{(L,\pm)} \end{bmatrix}, \tag{6.30}$$

the surface variation $[\delta H_{\mathbf{t}}^{(s)}]$ is found to be:

$$\begin{aligned}[\delta H_{\mathbf{t}}^{(s)}] = [\delta H_{\mathbf{t}}^{(s,\mathrm{nl})}] &+ \begin{bmatrix} \mathsf{G}_{\mathbf{t}}^{(A,+)} & \mathsf{G}_{\mathbf{t}}^{(B,-)} \end{bmatrix} \begin{bmatrix} \mathsf{V}^{(A,+)}(z_s) & 0 \\ 0 & \mathsf{V}^{(B,-)}(z_s) \end{bmatrix} \begin{pmatrix} \mathbf{a}^+ \\ \mathbf{b}^- \end{pmatrix} \\ &+ \begin{bmatrix} \mathsf{G}_{\mathbf{t}}^{(A,-)} & \mathsf{G}_{\mathbf{t}}^{(B,+)} \end{bmatrix} \begin{bmatrix} \mathsf{V}^{(A,-)}(z_s) & 0 \\ 0 & \mathsf{V}^{(B,+)}(z_s) \end{bmatrix} \begin{pmatrix} \mathbf{a}^- \\ \mathbf{b}^+ \end{pmatrix}. \end{aligned} \tag{6.31}$$

Inserting this equation and Eq. (6.15) into Eq. (6.28), the following matrix relation is derived:

$$\mathsf{L}^{AB}\begin{pmatrix} \mathbf{a}^+ \\ \mathbf{b}^- \end{pmatrix} = \begin{pmatrix} 0 \\ [\delta H_\beta^{(s,\mathrm{nl})}] \end{pmatrix} + \mathsf{R}^{AB}\begin{pmatrix} \mathbf{b}^+ \\ \mathbf{a}^- \end{pmatrix}, \tag{6.32}$$

where

$$\begin{aligned} \mathsf{L}^{AB} &= \begin{bmatrix} \mathsf{E}_{\mathbf{t}}^{(A,+)} & -\mathsf{E}_{\mathbf{t}}^{(B,-)} \\ \mathsf{H}_{\mathbf{t}}^{(A,+)} + \mathsf{G}_{\mathbf{t}}^{(A,+)} & -\mathsf{H}_{\mathbf{t}}^{(B,-)} - \mathsf{G}_{\mathbf{t}}^{(B,-)} \end{bmatrix} \begin{bmatrix} \mathsf{V}^{(A,+)}(z_s) & 0 \\ 0 & \mathsf{V}^{(B,-)}(z_s) \end{bmatrix}, \\ \mathsf{R}^{AB} &= \begin{bmatrix} \mathsf{E}_{\mathbf{t}}^{(B,+)} & -\mathsf{E}_{\mathbf{t}}^{(A,-)} \\ \mathsf{H}_{\mathbf{t}}^{(B,+)} + \mathsf{G}_{\mathbf{t}}^{(B,+)} & -\mathsf{H}_{\mathbf{t}}^{(A,-)} - \mathsf{G}_{\mathbf{t}}^{(A,-)} \end{bmatrix} \begin{bmatrix} \mathsf{V}^{(B,+)}(z_s) & 0 \\ 0 & \mathsf{V}^{(A,-)}(z_s) \end{bmatrix}. \end{aligned}$$

Here, $\mathsf{E}_{\mathbf{t}}^{(L,\pm)} = \left[\mathsf{E}_x^{(L,\pm)}; \mathsf{E}_y^{(L,\pm)}\right]$ denotes the $2N_0 \times 2N_0$ matrix of all tangential Fourier components of the modes in layer L. Note that the vectors of Fourier coefficient have already been arranged in a way suitable for the $\mathcal{S}$-matrix formalism, which is explained in what follows.

By multiplying from the left both sides of Eq. (6.32) with the inverse of the matrix L^{AB}, the Fourier coefficients of the outgoing modes, $\mathbf{a}^+$ and $\mathbf{b}^-$, can be determined in terms of the coefficients of the incoming modes, $\mathbf{a}^-$ and $\mathbf{b}^+$, and the nonlinear surface current:

$$\begin{aligned} \begin{pmatrix} \mathbf{a}^+ \\ \mathbf{b}^- \end{pmatrix} &= \left(\mathsf{L}^{AB}\right)^{-1}\left\{\begin{pmatrix} 0 \\ [\delta H_{\mathbf{t}}^{(s,\mathrm{nl})}] \end{pmatrix} + \mathsf{R}^{AB}\begin{pmatrix} \mathbf{b}^+ \\ \mathbf{a}^- \end{pmatrix}\right\} \\ &= \begin{pmatrix} \tilde{\mathbf{a}}^+ \\ \tilde{\mathbf{b}}^- \end{pmatrix} + \mathsf{S}^{AB}\begin{pmatrix} \mathbf{b}^+ \\ \mathbf{a}^- \end{pmatrix} = \begin{pmatrix} \tilde{\mathbf{a}}^+ \\ \tilde{\mathbf{b}}^- \end{pmatrix} + \begin{pmatrix} \overline{\mathbf{a}}^+ \\ \overline{\mathbf{b}}^- \end{pmatrix}, \end{aligned} \tag{6.33}$$

where $\mathsf{S}^{AB} = \left(\mathsf{L}^{AB}\right)^{-1}\mathsf{R}^{AB}$ is the scattering matrix ($\mathcal{S}$-matrix) of the interface system.

This equation shows that the outgoing coefficients are comprised of two different parts: $\overline{\mathbf{a}}^{+}$ and $\overline{\mathbf{b}}^{-}$ are the contributions to the total coefficients, $\mathbf{a}^{+}$ and $\mathbf{b}^{-}$, respectively, given by linear diffraction at the interface of the incident modes described by the coefficients $\mathbf{a}^{-}$ and $\mathbf{b}^{+}$, whereas $\tilde{\mathbf{a}}^{+}$ and $\tilde{\mathbf{b}}^{-}$ are the contributions of the nonlinear surface current to the total coefficients $\mathbf{a}^{+}$ and $\mathbf{b}^{-}$, respectively, and only enter if the generated frequency $\omega = \omega_0$ is considered.

If the considered structure only consists of cover, substrate, and one periodically patterned 2DM sheet, Eq. (6.33) is sufficient to fully describe its optical response. In multilayer structures, however, the contributions of the $\mathcal{S}$-matrices of different layers and interfaces have to be properly incorporated. For each of the pump frequencies, the near and far-field coefficients can be determined by the same $\mathcal{S}$-matrix formalism as described in Section 3.5.2 and Section 5.3.3, respectively, where the modified $\mathcal{S}$-matrix Eq. (6.33) is used for all 2DM sheets between two periodic bulk layers.

Thus, consider three layers, A, B, and S, with coefficients $\mathbf{a}^{\pm}$, $\mathbf{b}^{\pm}$, and $\mathbf{s}^{\pm}$, respectively, depicted in Fig. 6.6(b). Here, it is assumed that the $\mathcal{S}$-matrix, which connects the mode coefficients in layers B and S, is known and full fills the relation

$$\begin{pmatrix} \mathbf{b}^{+} \\ \mathbf{s}^{-} \end{pmatrix} = \mathsf{S}^{BS} \begin{pmatrix} \mathbf{b}^{-} \\ \mathbf{s}^{+} \end{pmatrix} + \begin{pmatrix} \tilde{\mathbf{b}}^{+} \\ \tilde{\mathbf{s}}^{-} \end{pmatrix}. \tag{6.34}$$

Combining this equation and Eq. (6.33), one determines the scattering matrix relation between the top and bottom coefficients, $\mathbf{a}^{\pm}$ and $\mathbf{s}^{\pm}$, respectively, by eliminating the coefficients $\mathbf{b}^{\pm}$:

$$\begin{pmatrix} \mathbf{a}^{+} \\ \mathbf{s}^{-} \end{pmatrix} = \mathsf{S}^{ABS} \begin{pmatrix} \mathbf{s}^{+} \\ \mathbf{a}^{-} \end{pmatrix} + \mathsf{T}^{ABC} \begin{pmatrix} \tilde{\mathbf{b}}^{+} \\ \tilde{\mathbf{b}}^{-} \end{pmatrix} + \begin{pmatrix} \tilde{\mathbf{a}}^{+} \\ \tilde{\mathbf{s}}^{-} \end{pmatrix}, \tag{6.35}$$

where the combined $\mathcal{S}$-matrix, $\mathsf{S}^{ABS} = \mathsf{S}^{AB} \otimes \mathsf{S}^{BC}$, is given by the Redheffer star-product Eq. (3.66). The four $N_0 \times N_0$ sub-blocks of the combined matrix T^{ABS} are expressed as:

$$\mathsf{T}_{11}^{ABS} = \mathsf{S}_{12}^{AB} \left(\mathsf{I} - \mathsf{S}_{11}^{BC} \mathsf{S}_{22}^{AB} \right)^{-1}, \tag{6.36a}$$

$$\mathsf{T}_{12}^{ABS} = \mathsf{S}_{12}^{AB} \left(\mathsf{I} - \mathsf{S}_{11}^{BS} \mathsf{S}_{22}^{AB} \right)^{-1} \mathsf{S}_{11}^{BS}, \tag{6.36b}$$

$$\mathsf{T}_{21}^{ABS} = \mathsf{S}_{21}^{BS} \left(\mathsf{I} - \mathsf{S}_{22}^{AB} \mathsf{S}_{11}^{BS} \right)^{-1} \mathsf{S}_{22}^{AB}, \tag{6.36c}$$

$$\mathsf{T}_{22}^{ABS} = \mathsf{S}_{21}^{BS} \left(\mathsf{I} - \mathsf{S}_{22}^{AB} \mathsf{S}_{11}^{BS} \right)^{-1}. \tag{6.36d}$$

These are found by straightforward matrix calculations, where one of the intermediate steps

yields the coefficients of the middle layer, B:

$$\mathbf{b}^- = \left(\mathsf{I} - \mathsf{S}_{22}^{AB}\mathsf{S}_{11}^{BS}\right)^{-1} \times \left[\mathsf{S}_{21}^{AB}\mathbf{a}^- + \mathsf{S}_{22}^{AB}\mathsf{S}_{12}^{BS}\mathbf{s}^+ + \tilde{\mathbf{b}}^- + \mathsf{S}_{22}^{AB}\tilde{\mathbf{b}}^+\right], \tag{6.37a}$$

$$\mathbf{b}^+ = \left(\mathsf{I} - \mathsf{S}_{11}^{BS}\mathsf{S}_{22}^{AB}\right)^{-1} \times \left[\mathsf{S}_{11}^{BS}\mathsf{S}_{21}^{AB}\mathbf{a}^- + \mathsf{S}_{12}^{BS}\mathbf{s}^+ + \mathsf{S}_{11}^{BS}\tilde{\mathbf{b}}^- + \tilde{\mathbf{b}}^+\right] \tag{6.37b}$$

expressed solely in terms of incoming and known coefficients $\mathbf{a}^-$, $\mathbf{s}^+$, and $\tilde{\mathbf{c}}^\pm$.

Note that the term

$$\mathsf{T}^{ABS}\begin{pmatrix}\tilde{\mathbf{b}}^+\\ \tilde{\mathbf{b}}^-\end{pmatrix} + \begin{pmatrix}\tilde{\mathbf{a}}^+\\ \tilde{\mathbf{s}}^-\end{pmatrix} =: \begin{pmatrix}\tilde{\mathbf{a}}_{\text{eff}}^+\\ \tilde{\mathbf{s}}_{\text{eff}}^-\end{pmatrix} \tag{6.38}$$

in Eq. (6.35) can be interpreted as the effective coefficients of the modes that are radiated by the combined multilayer-interface system, ABS. More specifically, this term accounts for the linear propagation of the internally radiated modes at the generated frequency, with coefficients $\tilde{\mathbf{b}}^\pm$, in bulk layer B and it accounts for linear optical interaction (reflection, transmission, and absorption) with the 2DM located at the AB-interface or the BS-layer-interface system.

Equipped with these matrix-vector relations, the calculation of all solution coefficients at the generated frequency can be completed. In order to evaluate the mode coefficients $\mathbf{e}^\pm$ in the evaluation layer E, one calculates the combined $\mathcal{S}$-matrix, S^C, of all layers and interfaces above the evaluation layer E, and the combined $\mathcal{S}$-matrix, S^M, of all interfaces and layers between the evaluation layer E and up to but excluding the 2DM sheet. This is depicted in Fig. 6.6(c). Note that this procedure allows for S^C, S^M or S^{BS} to be equal to $\mathsf{I}^\otimes = [0, \mathsf{I}; \mathsf{I}, 0]$, i.e. the evaluation layer can be any layer above the nonlinear 2DM-sheet, which itself can be located at any interface, including just above the substrate.

Under these circumstances, the governing $\mathcal{S}$-matrix relations read as follows:

$$\begin{pmatrix}\mathbf{c}^+\\ \mathbf{e}^-\end{pmatrix} = \mathsf{S}^C\begin{pmatrix}\mathbf{c}^-\\ \mathbf{e}^+\end{pmatrix}, \tag{6.39a}$$

$$\begin{pmatrix}\mathbf{e}^+\\ \mathbf{a}^-\end{pmatrix} = \mathsf{S}^M\begin{pmatrix}\mathbf{e}^-\\ \mathbf{a}^+\end{pmatrix}, \tag{6.39b}$$

$$\begin{pmatrix}\mathbf{a}^+\\ \mathbf{s}^-\end{pmatrix} = \mathsf{S}^{ABS}\begin{pmatrix}\mathbf{a}^-\\ \mathbf{s}^+\end{pmatrix} + \begin{pmatrix}\tilde{\mathbf{a}}_{\text{eff}}^+\\ \tilde{\mathbf{s}}_{\text{eff}}^-\end{pmatrix}. \tag{6.39c}$$

By applying Eq. (6.35) to the matrix relations for S^M and S^{ABS}, one obtains $\mathsf{S}^{MABS} = \mathsf{S}^M \otimes \mathsf{S}^{ABS}$ and effective affine coefficients, $\tilde{\mathbf{e}}_{\text{eff}}^+ := \mathsf{T}_{11}^{MABS}\tilde{\mathbf{a}}_{\text{eff}}^+$ and $\tilde{\mathbf{s}}_{\text{eff}}^- := \mathsf{T}_{21}^{MABS}\tilde{\mathbf{a}}_{\text{eff}}^+ + \tilde{\mathbf{s}}_{\text{eff}}^-$, as per Eq. (6.38).

The final constellation of the remaining two $\mathcal{S}$-matrices, S^C and S^{MABS}, and coef-

ficients is depicted in Fig. 6.6(d). This situation is analogous to the initial system of $\mathcal{S}$-matrices S^{AB} (Eq. (6.33)) and S^{BS} Eq. (6.34). Hence, it allows the calculation of the evaluation coefficients $\mathbf{e}^{\pm}$ by means of Eq. (6.37), with the following replacements $\mathbf{b}^{\pm} \to \mathbf{e}^{\pm}$, $\mathsf{S}^{AB} \to \mathsf{S}^{C}$, $\mathsf{S}^{BS} \to \mathsf{S}^{MABS}$, $\mathbf{a}^{-} \to 0$, $\mathbf{s}^{+} \to 0$, $\tilde{\mathbf{b}}^{-} \to 0$, and $\tilde{\mathbf{a}}^{+} \to \tilde{\mathbf{e}}^{+}_{\text{eff}}$, yielding

$$\mathbf{e}^{-} = \left(\mathsf{I} - \mathsf{S}^{C}_{22} S^{MABS}_{11}\right)^{-1} \mathsf{S}^{C}_{22} \tilde{\mathbf{e}}^{+}_{\text{eff}}, \tag{6.40}$$

$$\mathbf{e}^{+} = \left(\mathsf{I} - \mathsf{S}^{MABS}_{11} \mathsf{S}^{C}_{22}\right)^{-1} \tilde{\mathbf{e}}^{+}_{\text{eff}} \tag{6.41}$$

The treatment of an evaluation layer below the nonlinear layer can be performed in a similar manner. Moreover, if the structure contains more than one nonlinear 2DM sheet, the algorithm just described can be repeated independently for each interface where nonlinear 2DM is located and then sum the individually obtained solution coefficients. It should be noted that in addition to the nonlinear optical response of the 2DMs the linear scattering effects at the interfaces are naturally incorporated in our algorithm. This overall approach to linear and nonlinear light scattering in layered photonic structures containing 2DMs is possible because of the affine linearity of the total system in the undepleted pump approximation.

The mathematical formulation of the RCWA for 2D materials was implemented in a separate MATLAB [41] program and is not integrated in OmniSim RCWA. However, similar software design principles as were described in Chapter 5 have been applied for this secondary implementation, and special care was taken in order to ensure the correctness and validation of this software tool.

6.5 Validation of the numerical method and convergence analysis

This section is concerned with the validation of the presented numerical method by considering a series of test cases and analysis of its convergence characteristics. Both 1D- and 2D-periodic structures made of graphene with cubic nonlinearity are considered. No separate test case for the TMDC monolayers is shown here, because they are less challenging than graphene from a computational pint of view: TMDCs are poorly conductive and do not influence the EM fields as strongly as graphene does. Both EM near- and far-field quantities have to be considered to assess the stability and convergence of a modal method as was found in Ref. [40]. To characterise the far-field, the optical absorption A in the grating is used at the FF, as its numerical value is given by $A = (I - R - T)/I$, where I, R, and T denote the incident, reflected, and transmitted power, respectively. At the TH, the total radiation, $R' + T'$, is chosen as a far-field quantity suitable to validate the method, where R' and T' denote the intensity at the TH radiated in direction of reflection and transmission,

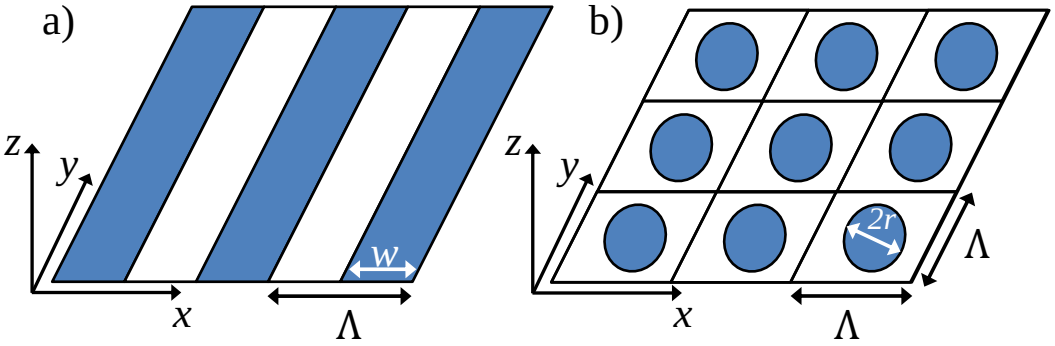

Figure 6.8: Generic structures for convergence analysis: a) 1D periodic array of graphene ribbons with period, Λ, and width, w. b) 2D array circular graphene disks with radius, r, and periods $\Lambda_1 = \Lambda_2 = \Lambda$.

respectively.

6.5.1 One-dimensional binary graphene gratings

As a first example of a periodic diffraction grating containing nonlinear 2DMs, consider the 1D periodic array of graphene ribbons depicted in Fig. 6.8(a), sandwiched in-between homogeneous cover and substrate materials with electric permittivity, $\epsilon_c = 3$ and $\epsilon_s = 4$, respectively. The period of the grating is $\Lambda = 8\,\mu\mathrm{m}$ and the spacing between adjacent ribbons is $\Lambda - w = \Lambda/2 = 4\,\mu\mathrm{m}$. The incident light is normally impinging onto this binary graphene grating and is TM-polarised.

Three different variations of the proposed algorithm are compared here to showcase the necessity and benefit of the model with additional conductivity Eq. (6.22). *i)* Use a nonzero additional conductivity $\sigma_{s,\mathrm{add}} = -i\eta|\sigma_s(\omega)|$ and apply the correct Fourier factorisation Eq. (6.24). *ii)* Do not use the additional conductivity ($\sigma_{s,\mathrm{add}} = 0$) and only use the product factorisation rule for Eq. (6.17), i.e. use an incorrect factorisation rule. *iii)* Assume a finite thickness of the 2DM, $h_{\mathrm{eff}} = 0.33\,\mathrm{nm}$, i.e. model the 2DM-sheet it a periodic bulk layer with relative permittivity $\epsilon_r = 1 + i\sigma_s/(\epsilon_0\omega h_{\mathrm{eff}})$. This can be done in any RCWA implementation. It should be stressed, however, that this is computationally more costly as it involves the determination of the RCWA modes in the bulk layer representing the 2DM. Additionally, to validate the results, numerical results from Ref. [8] are included as a comparison.

The linear absorption spectra calculated using a moderate number of harmonics, $N = 100$, are show in Fig. 6.9. The device absorption presents a broad resonance peak with maximum of about $18.5\,\%$ at the wavelength, $\lambda = 80\,\mu\mathrm{m}$, with negligible absorption being observed at both short and long wavelengths. The results obtained using the algorithm with added conductivity with $\eta = 10^{-5}$ shows very good agreement with those found using conventional bulk RCWA calculations and with the results taken from Ref. [8]. The model without additional conductivity slightly overestimates the absorption in the region

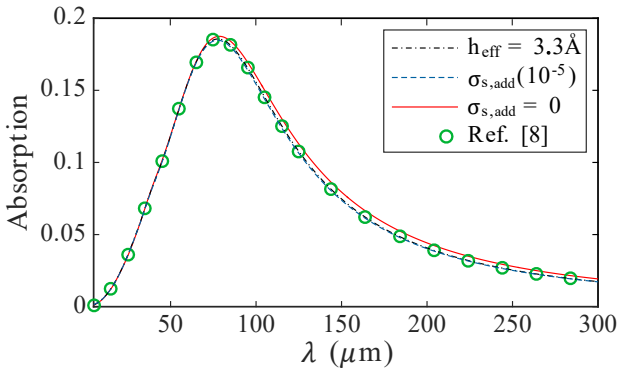

Figure 6.9: Linear absorption spectrum obtained by using the model without additional conductivity (solid line), by using the model with $\sigma_{s,\mathrm{add}}(10^{-5})$ (dashed line), using conventional RCWA with $h_{\mathrm{eff}} = 0.33\,\mathrm{nm}$ (dashed-dotted line) and results from Ref. [8] (green circles).

$\lambda \gtrsim 70\,\mu\mathrm{m}$.

Keeping in mind that the additional sheet conductance, $\sigma_{s,\mathrm{add}}(\eta) = -i\eta|\sigma^s(\omega)|$, has been introduced somewhat artificially to facilitate the application of the correct (inverse) factorisation rule, the influence of $\sigma_{s,\mathrm{add}}(\eta)$ on the accuracy of the computed results needs to be carefully investigated. The asymptotic behaviour of both near- and far-field physical quantities are suitable tools for performing this analysis. Thus, the dependence of the absorption at $\lambda = 18.5\,\mu\mathrm{m}$ on N and η, was determined and is depicted in Fig. 6.10(a). Among other things, this figure shows that the convergence with respect to N is faster for larger η. This dependence is not surprising, because the interface containing the structured graphene sheet becomes more optically homogeneous as η increases, i.e. the description of $1/\tilde{\sigma}_s(x,y)$ as a Fourier series becomes more accurate. However, this does not prove the accuracy of the method just yet: $\sigma_{s,\mathrm{add}}$ was introduced at a purely mathematical level and as such it should vanish in physical diffraction gratings. The accuracy of the method is demon-

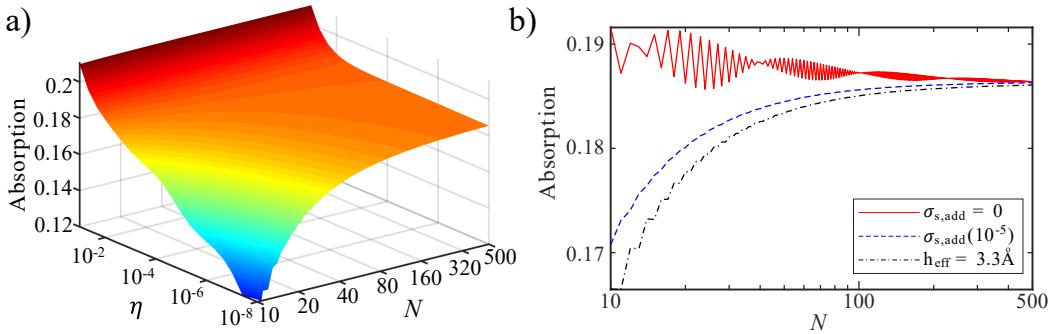

Figure 6.10: a) Convergence graph for increasing N and decreasing $\sigma_{s,\mathrm{add}}(\eta)$ using the model with additional conductivity. b) The dependence of absorption on N, calculated using a conventional RCWA (dashed-dotted line), the algorithm with $\sigma_{s,\mathrm{add}} = 0$ (solid line), and the algorithm with added conductivity, $\sigma_{s,\mathrm{add}} = \sigma_{s,\mathrm{add}}(10^{-5})$ (dashed line).

strated by the fact that as $\eta \to 0$, convergence of the absorption is reached; for example, as $\eta \to 0$ the maximum absorption converges to $A = 18.63\,\%$ for increasing value of N.

Figure 6.10(b) contrasts the convergence characteristics of the three algorithms that were just described.As it can be seen in this figure, all three approaches converge to the same value but, importantly, the convergence speed in the case of finite added conductivity is the fastest among the three cases. It is also instructive to remark that the slowest convergence is observed in the case of zero conductivity, an added drawback in this case being the oscillatory dependence of the absorption on N, which confirms that this formulation is incorrect as was theoretically argued in Section 6.4.3.

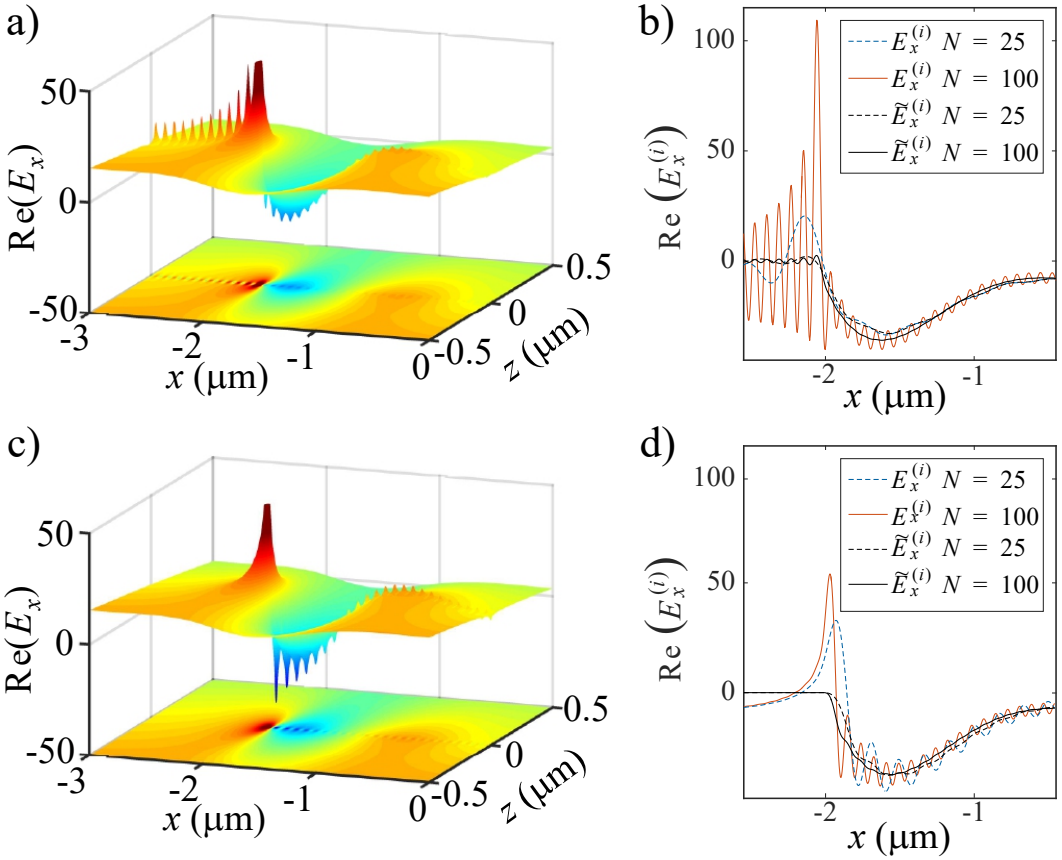

Figure 6.11: a), b) Electric field component E_x around the graphene ribbon calculated for $N = 100$ and the fields at the surface of the graphene ribbon corresponding to the improper (coloured lines) and correct (black lines) field evaluation, respectively. In both cases, $\sigma_{s,\mathrm{add}} = 0$. c), d) The same quantities as in a) and b), respectively, but determined for $\sigma_{s,\mathrm{add}} = \sigma_{s,\mathrm{add}}(10^{-5})$. Only a part of the left half of the unit cell, defined by $x \in [-4\mu\mathrm{m}, 0]$, is shown, as the results are symmetric in x and rather featureless far away from the graphene ribbon located at $|x| < 2\,\mu\mathrm{m}$.

The spatial profile of the electric near-field deserves special attention as well because it reveals new important features pertaining to the convergence of the numerical method. To illustrate this idea, the x-component of the electric field for the cases in which $\sigma_{s,\mathrm{add}} = 0$

and when the conductivity is finite, $\sigma_{s,\mathrm{add}} = \sigma_{s,\mathrm{add}}(10^{-5})$, are shown in Figs. 6.11(a) and 6.11(c), respectively. The operating wavelength is $\lambda = 33\,\mu\mathrm{m}$ and $N = 100$ harmonics in both cases. In these plots, the boundary of the graphene ribbon is located at $x = -2\,\mu\mathrm{m}$. It can be seen from these plots that without additional conductivity the field exhibits very strong, unphysical oscillations near the boundary of the graphene ribbons at $i.e. at z = z_s = 0$, which spread over the whole unit cell along the x-direction. The spatial frequency of these oscillations is equal to the highest spatial discretisation frequency, namely to $2\pi N/\Lambda$. In the case of finite $\sigma_{s,\mathrm{add}}$, on the other hand, there are no such spurious field oscillations outside graphene regions and only weak oscillations are seen inside graphene regions, as per Fig. 6.11(c) where $\eta = 10^{-5}$. In both cases, these field oscillations only occur very close to the interface where the 2DM is located and can merely be observed at distances $|z - z_s| \gtrapprox 5\,\mathrm{nm}$. These oscillations are due to the fact that the Fourier series decomposition does not resolve the singularity of the electric field at the edges of the graphene ribbons, as was discussed in Section 6.4.3.

One of the aims of the accurate near-field formulation introduced in Eq. (6.26) is to overcome this shortcoming of Fourier series decompositions and thus to allow the accurate field evaluation exactly at the location of the 2DM monolayers, i.e. at $z = z_s$. To illustrate how this is achieved here, Figs. 6.11(b) and 6.11(d) depict the interface field, $E_x^{(i)}(x, y, z = z_i)$, for the cases in which $\sigma_{s,\mathrm{add}} = 0$ and when the conductivity is finite, $\sigma_{s,\mathrm{add}} = \sigma_{s,\mathrm{add}}(10^{-5})$, respectively. Without using an added conductivity, the improper field evaluation $\mathscr{R}\left([E_x^{(i)}]\right)(x, y)$ of the surface field leads to a strongly oscillatory spatial dependence, which in addition changes significantly with N, as shown by the blue and red lines in Fig. 6.11(b) for $N = 25$ and $N = 100$, respectively. By contrast, the interface field obtained by using Eq. (6.26) with $[\![1/\tilde{\sigma}_s]\!]^{-1}$ replaced by $[\![\tilde{\sigma}_s]\!]$ displays hardly any oscillations, as per the black lines in this figure, yet it does not vanish at the edge of the graphene ribbon as required by the theoretical analysis presented in Section 6.4.3. These remaining oscillations are merely due to the incorrect factorisation rule and not due to the unresolved singularity of the field.

This description changes significantly if one uses a small, finite value for $\sigma_{s,\mathrm{add}}$, see Fig. 6.11(d). The improper evaluation using a finite value for $\sigma_{s,\mathrm{add}}$ yields a smooth field outside the graphene region ($|x| > 2\,\mu\mathrm{m}$), but still suffers from oscillations in the graphene region ($|x| \leq 2\,\mu\mathrm{m}$) and does not vanish at the boundary of the graphene ribbon ($|x| = 2\,\mu\mathrm{m}$), as required. Finally, the correct evaluation of the surface field and surface current given by Eqs. Eq. (6.26) vanishes at $|x| = 2\,\mu\mathrm{m}$, is free of spurious oscillations, and converges rapidly with N.

Because the nonlinear polarisations generated in the photonic structures here are determined by the optical near-field at the location of the 2DMs, it has to be stressed that an

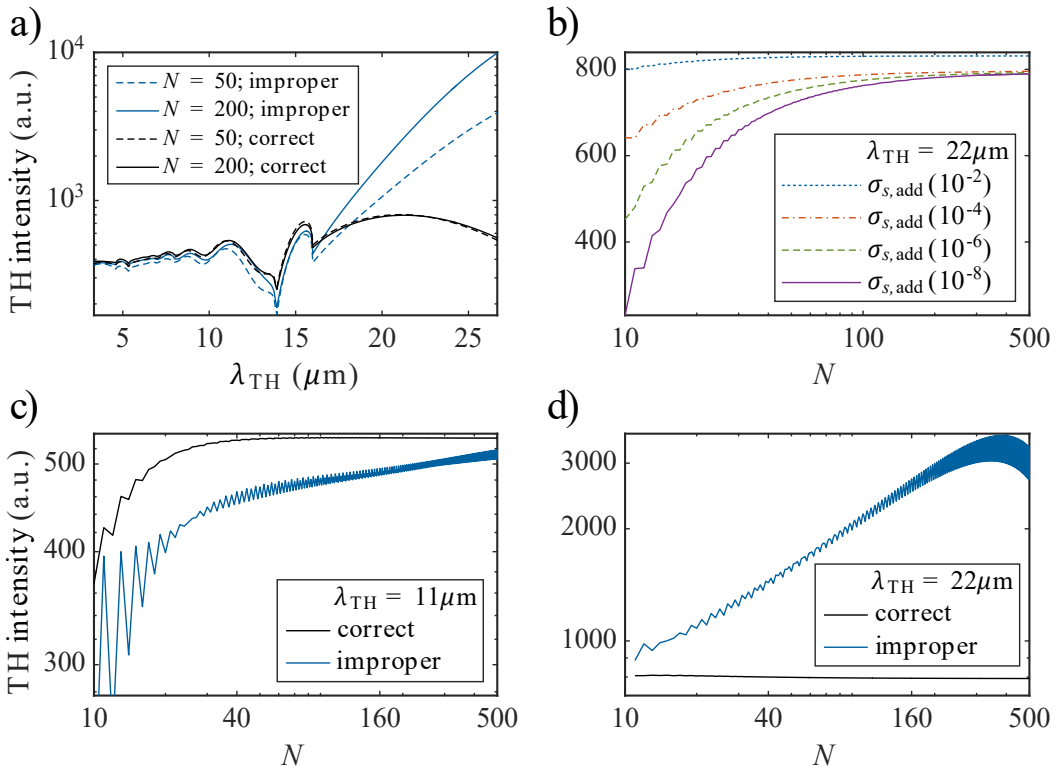

Figure 6.12: a) Nonlinear radiation spectrum for $N = 50$ (dashed lines) and $N = 200$ (solid lines) calculated using the improper field evaluation (blue lines) and correct field evaluation (black lines) at the FF. b) Intensity of TH radiation at $\lambda_{TH} = 22\,\mu m$ vs. N determined for different values of added conductivity $\sigma_{s,add}$. c), d) Intensity of TH radiation at $\lambda_{TH} = 11\,\mu m$ and $\lambda_{TH} = 22\,\mu m$, respectively, vs. N determined using improper field evaluation (blue lines) and correct field evaluation (black lines) at the FF. In both cases $\sigma_{s,add} = \sigma_{s,add}(10^{-5})$.

accurate evaluation of these near-fields is a crucial prerequisite to a rigorous description of the nonlinear optical response of these structures. As was just shown, this can be achieved by using the correct Fourier factorisation Eq. (6.24) in the model with finite conductivity and the correct field evaluation Eq. (6.26).

The total intensity of radiated TH is considered as a first quantity characterizing the nonlinear optical response of the graphene grating and the corresponding computational results are summarised in Fig. 6.12. Thus, Fig. 6.12(a) depicts the total radiated power at the TH, determined for the case of finite added conductivity, $\sigma_{s,add}(10^{-5})$, and number of harmonics, $N = 50$ and $N = 200$. The nonlinear source current given by Eq. (6.5) was calculated using either the improperly evaluated interface field, $\mathbf{E}^{(i)}$, or the correctly calculated interface field, $\tilde{\mathbf{E}}^{(i)}$. Interestingly, in the wavelength range $\lambda_{TH} < 15\,\mu m$, both calculation methods yield qualitatively similar spectra. For longer wavelengths, however, results differ considerably and in fact only the results obtained using the $\tilde{\mathbf{E}}^{(i)}$ field converge.

Before investigating in more detail this behaviour, consider first the convergence char-

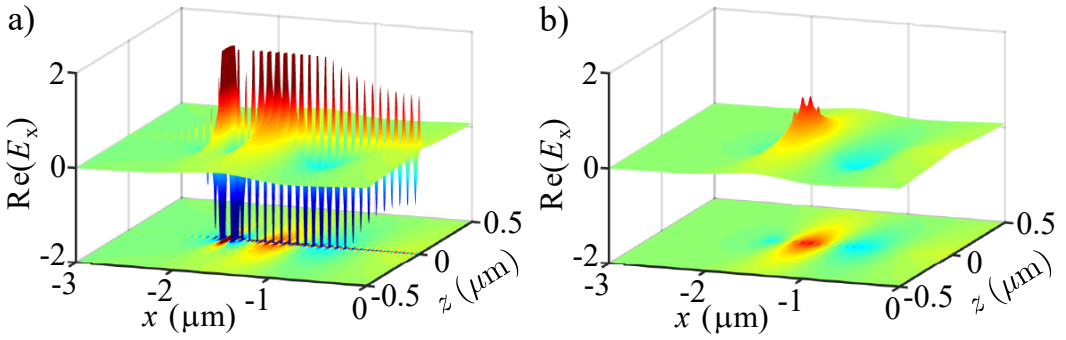

Figure 6.13: a) Third-harmonic near-field around a graphene ribbon determined using the improperly evaluated interface field at the FF, $E_x^{(i)}$, to calculate the nonlinear polarisation, $\mathbf{P}^{\mathrm{nl}}$. b) The same as in a) but using the correctly evaluated field at the FF, $\tilde{E}_x^{(i)}$, to compute $\mathbf{P}^{\mathrm{nl}}$.

acteristics of the far-field at the TH with respect to the value of the added conductivity, $\sigma_{s,\mathrm{add}}$. The main features of this dependence, illustrated by the data plotted in Fig. 6.12(b), which corresponds to $\lambda_{\mathrm{TH}} = 22\,\mu\mathrm{m}$, are similar to those observed in the case of linear calculations. More specifically, the larger $\sigma_{s,\mathrm{add}}$ is the faster self-convergence with respect to N is observed and the computational results converge for vanishingly small $\sigma_{s,\mathrm{add}}$.

The difference in convergence behaviour for increasing N of the two approaches used to evaluate the surface field is illustrated in Figs. 6.12(c) and 6.12(d), for two representative wavelengths, $\lambda_{\mathrm{TH}} = 11\,\mu\mathrm{m}$ and $\lambda_{\mathrm{TH}} = 22\,\mu\mathrm{m}$, respectively. These figures show that the correct field evaluation leads to rapid convergence at both wavelengths, whereas the improper approach yields slow and oscillatory convergence at $\lambda_{\mathrm{TH}} = 11\,\mu\mathrm{m}$ and completely fails to converge at $\lambda_{\mathrm{TH}} = 22\,\mu\mathrm{m}$.

The electromagnetic near-field, $\mathbf{E}^{(\Omega)}$, at the TH wavelength, $\lambda_{\mathrm{TH}} = 11\,\mu\mathrm{m}$, determined using the two algorithmic choices and $N = 100$ spatial harmonics, is plotted in Fig. 6.13. Thus, the TH E_x field derived from the nonlinear source polarisation, $\mathbf{P}^{\mathrm{nl}}(\mathbf{E}^{(i)})$, comprising the incorrectly evaluated field at the FF, $\mathbf{E}^{(i)}$, is completely swamped by unphysical oscillations, i.e. numerical artefacts, at the interface where the graphene ribbon is located, as per Fig. 6.13(a). On the other hand, the x-component of the TH electric field, $E_x^{(\Omega)}$, which has the nonlinear polarisation $\mathbf{P}^{\mathrm{nl}}(\tilde{\mathbf{E}}^{(i)})$ as its source, shows almost no oscillatory behaviour and thus has the expected spatial profile, as shown in Fig. 6.13(b). Only further away from the interface, some agreement between the data in Figs. 6.13(a) and 6.13(b) can be observed.

The faster convergence, the less spurious near-fields at both FF and TH, combined with the expected electric field dependence at the graphene boundaries allow us to conclude that the properly evaluated field, $\tilde{\mathbf{E}}^{(i)}$, at the FF yields the correct results at the TH, whereas the TH optical response obtained using the improperly evaluated field, $\mathbf{E}^{(i)}$, at the FF is plagued

by numerical artefacts and unphysical behaviour.

6.5.2 Two-dimensional graphene diffraction gratings

Let us now consider 2D diffraction gratings that contain 2DMs, namely the periodically arranged graphene disks depicted in Fig. 6.8(b). Let the two periods be the same, $\Lambda_1 = \Lambda_2 = \Lambda = 250\,\mathrm{nm}$ and the diameter of the graphene disks be $2r = 175\,\mathrm{nm} = 0.7\Lambda$. For the sake of simplicity the incoming light is normally impinging onto the grating, the cover and substrate media being air ($\epsilon_c = 1$) and glass ($\epsilon_s = 2.0852$), respectively.

The absorption spectrum for x-polarised light, calculated for $N = 10$, 15, 20, and 25 harmonics, is presented in Fig. 6.14(a). All results are obtained using the finite conductivity model defined by Eq. (6.22), the scaling parameter being $\eta = 10^{-3}$. The zero-conductivity model or the improper interface field evaluation yield highly oscillatory near-field profiles at FF and TH and fail to deliver converging TH far-field results. Thus, it can be seen that the spectra calculated using different values of N agree well and exhibit similar features. In particular, all spectra show a series of resonances whose amplitude and spectral width increase with the wavelength. Figure 6.14(b) depicts the dependence of the intensity of the total radiated TH from the array of graphene disks, R'_N, on the number of harmonics, N. The nonlinear radiation spectra corresponding to the largest values $N = 25$ and $N = 20$

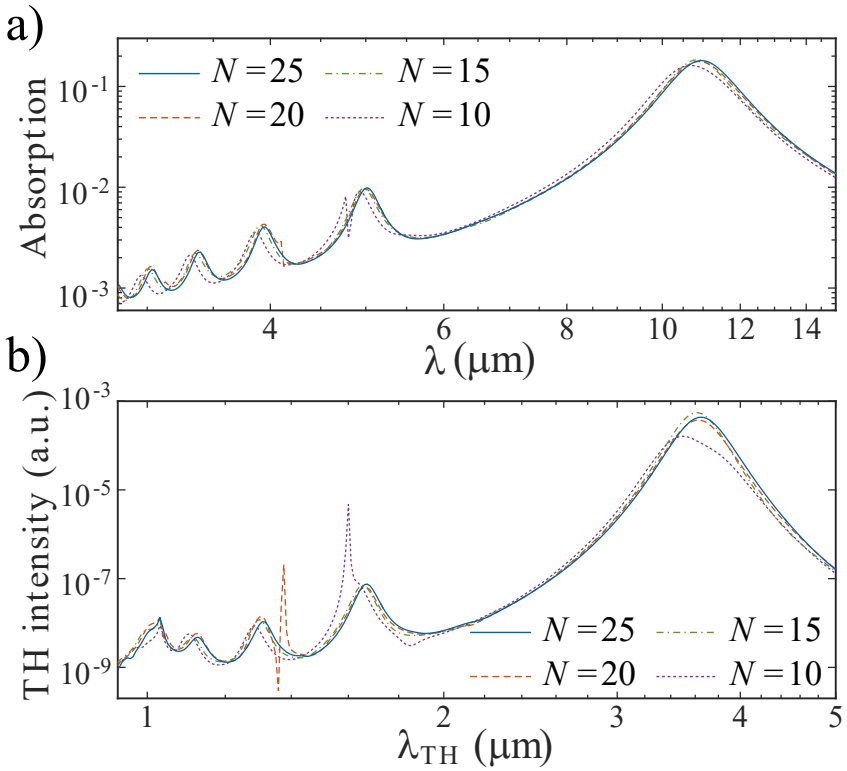

Figure 6.14: a), b) Linear absorption and intensity of TH radiation spectra calculated using a finite added conductivity, $\sigma_{s,\mathrm{add}} = \sigma_{s,\mathrm{add}}(10^{-3})$, and number of harmonics $N = 10$, 15, 20, and 25.

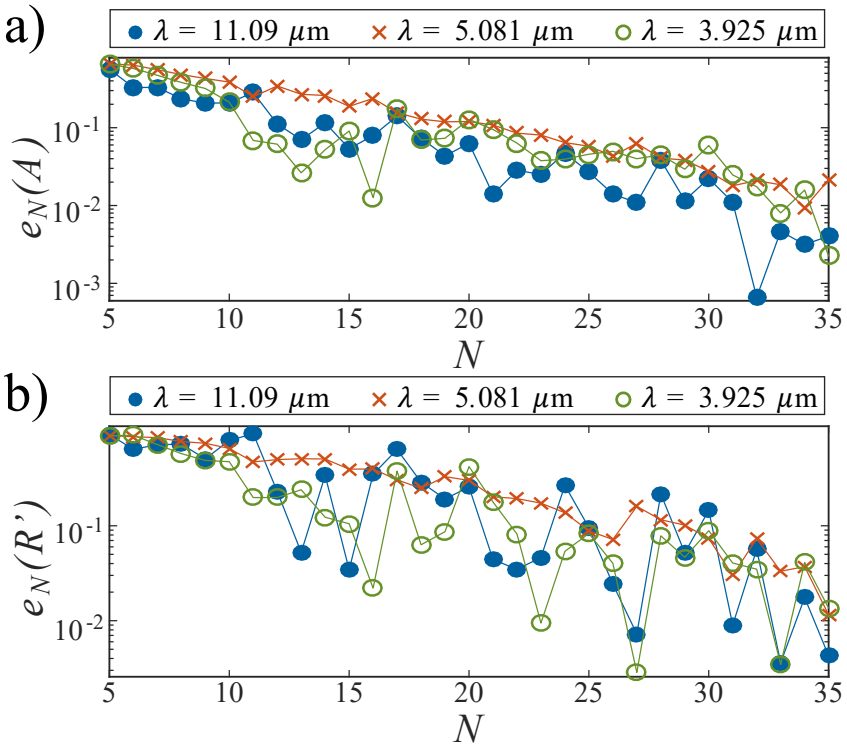

Figure 6.15: a), b) Relative self-error for linear absorption and intensity of TH radiation, respectively, determined at the plasmon resonance wavelengths $\lambda = 11.09\,\mu\text{m}$ (dots), $\lambda = 5.081\,\mu\text{m}$ (crosses) and $\lambda = 3.925\,\mu\text{m}$ (circles)

already show good agreement, which suggests that convergence has been achieved.

The self-convergence of linear calculations is investigated in more detail, the main conclusions being summarised in Fig. 6.15(a). This investigation is performed at the wavelengths of the first three absorption resonances, that is $\lambda = 11.09\,\mu\text{m}$, $\lambda = 5.081\,\mu\text{m}$, and $\lambda = 3.925\,\mu\text{m}$. Setting as converged values the results obtained by using a large number of harmonics, $N = 40$, specifically $A^* = A_{40}$, the relative self-error corresponding to a number of harmonics, N, is then defined as:

$$e_N(A) = \frac{A_N - A^*}{A^*}. \tag{6.42}$$

This self-error function can be used as a reliable measure of the convergence of the method as long as the value of N for which the reference absorption is defined is chosen to be sufficiently large.

At all three resonance wavelengths, relative errors of $e_N(A) \approx 1\,\%$ are achieved for $N \geq 35$. It should be noted that in terms of computational effort, using $N_{2D} = 35$ in 2D simulations is comparable to using $N_{1D} = 2520$ in 1D simulations. Moreover, the accuracy of the 2D simulation with $N_{2D} = 35$ is still correlated to and limited by the highest spatial frequency, $2\pi N_{2D}/\Lambda$.

To asses the convergence of the TH simulations more rigorously, the relative error, $e_N(R')$, of the TH radiation intensity, R'_N, for a given number of harmonics N, which is defined similarly to $e_N(A)$ in Eq. (6.42) with $R'^* = R'_{40}$, is depicted in Fig. 6.15(b). A relative self error of $e_N(R') \approx 2\%$ can be achieved at $N = 35$ for all three resonance wavelengths. Note also that the intensity of the TH radiation varies over six orders of magnitude and has maxima at the locations of the spectral resonances of the linear absorption, as per Fig. 6.14(a).

Graphene is a lossy conductor in the spectral range considered and therefore allows the excitation of surface waves [4, 6, 19]. This is the case with each of the absorption maxima seen in the linear spectrum, as illustrated by Figs. 6.16(a)–6.16(c). Thus, they show the dominant field component, $|E_x|$, of the linear electric field at the first three resonance wavelengths, $\lambda = 11.09\,\mu\mathrm{m}$ in Fig. 6.16(a), $\lambda = 5.081\,\mu\mathrm{m}$ in Fig. 6.16(b), and $\lambda = 3.925\,\mu\mathrm{m}$ in Fig. 6.16(c). These field profiles exhibit distinct mode shapes with one, three and five maxima, which demonstrates that the corresponding resonances represent the first three plasmon modes of the graphene disks.

Strongly enhanced and largely confined optical field resulting from the excitation of localised plasmon modes gives rise to enhanced THG. This fact is supported by the field profiles plotted in the bottom panels of Fig. 6.16, where the dominant electric field component at the TH, $|E_x|$, is shown. Indeed, regions of strong field enhancement can be seen, with field profiles with three, seven, and eleven maxima being observed at $\lambda_{\mathrm{TH}} = 11.09\,\mu\mathrm{m}/3$ in Fig. 6.16(d), $\lambda_{\mathrm{TH}} = 5.081\,\mu\mathrm{m}/3$ in Fig. 6.16(e), and $\lambda_{\mathrm{TH}} = 3.925\,\mu\mathrm{m}/3$ in Fig. 6.16(f), respectively.

6.6 Diffraction in nonlinear gratings containing 2D materials

In this section the linear and nonlinear optical response of diffraction gratings incorporating TMDC monolayer materials or graphene is investigated and the application of the inhomogeneous $\mathcal{S}$-matrix formulation introduced in Section 6.4.4 is illustrated in several specific cases.

Throughout this section, the intensity of the incident beam is chosen to be $I_0 = 10^{12}\,\mathrm{W\,m^{-2}}$, which is a moderately high peak intensity generated by a pulsed laser. Changing the incident intensity in the undepleted pump approximation does not alter the numerical results at any of the incident frequencies, $n = 1, \ldots, N_F$; however, it substantially changes the magnitude of the electromagnetic field at the generated frequency, $n = 0$. More specifically, for SHG and THG the intensity of the generated waves behaves as $I_{\mathrm{SH}} \propto I_0^2$ and $I_{\mathrm{TH}} \propto I_0^3$, respectively, and as such they can increase to significant values. A problem that might arise in the undepleted pump approximation is that if the intensity of the generated waves becomes comparable to the intensity of the incident wave, the use of this approximation would become questionable. This is not the case in most practical situations and

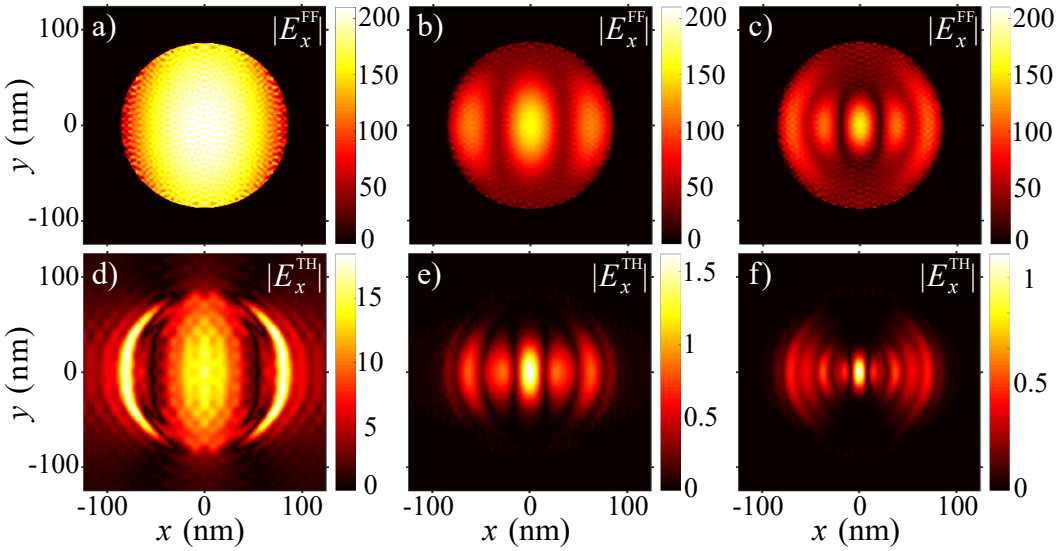

Figure 6.16: Dominant electric field component, $|E_x|$, at the FF (top panels) and TH (bottom panels) at the surface of a graphene disk for the resonance wavelengths, from left to right, $\lambda_{FF} = 11.09\,\mu\text{m}$, $\lambda_{FF} = 5.081\,\mu\text{m}$, and $\lambda_{FF} = 3.925\,\mu\text{m}$. The number of harmonics used in the simulations was $N = 40$ at the FF and $N = 30$ at the TH.

certainly not the case here, as can be seen by the intensity of generated optical fields in the examples hereafter.

6.6.1 SHG from TMDC monolayer ribbons

To begin with, a 1D binary TMDC grating placed on top of a glass substrate with $\epsilon_s = 1.44$ is considered. Its period is $\Lambda = 100\,\text{nm}$ and the filling factor is 0.9. An x-polarised plane wave is normally incident onto the grating and a spectral range of $0.4\,\mu\text{m}$ to $4\,\mu\text{m}$ for the incident wavelength λ is considered. The computations were performed using $N = 200$ harmonics and an added sheet conductance of $\sigma_{s,\text{add}} = -i10^{-5}|\sigma_s(\omega)|$, the results being presented in Fig. 6.17. The convergence of these calculations has been assured as diligently as for the binary graphene diffraction gratings discussed in Section 6.5.1.

The plots presented in Fig. 6.17(a) reveal that the linear absorption spectra are primarily determined by the linear material properties $\sigma_s(\omega)$ of the TMDC materials. The absence of any additional resonant features in the spectra has two main reasons: the dielectric nature of the TMDCs does not allow the formation of plasmons, as in the case of graphene, whereas the vanishingly small thickness of the TMDC monolayers precludes the existence of geometric, Mie-type resonances. These explanations are supported also by the fact that linear and nonlinear spectra are qualitatively similar if one varies the grating period and the feature size of the TMDC ribbons.

In order to obtain the SH radiation of the 2DM grating, the nonlinear conductivity, $\boldsymbol{\sigma}^{(2)}$, from Section 6.3.2 was used. Since no values of $\boldsymbol{\sigma}^{(2)}$ for $MoSe_2$ were available, the nonlinear optical response of the grating was determined only in the case of three TMDC

materials, WS_2, MoS_2, and WSe_2. As the calculations showed that there is no resonant field enhancement at the TMDC ribbons at the FF, no strong enhancement of generated SH is expected. The nonlinear SH radiation spectra presented in Fig. 6.17(b) suggest that indeed the intensity of radiated SH closely follows the magnitude of $\boldsymbol{\sigma}^{(2)}$, as can be found by comparison with the plots in Fig. 6.5(b). The maximal intensity of generated SH is $I_{\mathrm{SH}} = 2 \times 10^{-9} I_0$, rendering valid the undepleted pump approximation.

6.6.2 Enhancement of the nonlinear efficiency for TMDC monolayers on a slab waveguide

Since TMDC monolayers themselves do not possess optical modes, a different approach to achieve enhanced nonlinear optical interactions in these 2D materials is used. Thus, a TMDC monolayer is combined with a bulk structure that possesses waveguide modes whose excitation leads to strong local field enhancement. A very effective structure for this purpose is a periodically patterned slab waveguide, covered by a TMDC monolayer, as depicted in Fig. 6.18(a). Dielectric slab waveguides exhibit very narrow spectral features, due to the resonant excitation of guiding slab modes, a phenomenon that can be used to enhance linear and nonlinear optical response of specific devices [42–44].

The waveguide structure under consideration consists of a slab of height, h, placed between a monolayer WS_2 and substrate with relative permittivity $\epsilon_s = 1.46^2$ (index of refraction, $n_s = 1.46$). The 2DM monolayer is adjacent to the cover region, which is assumed to be vacuum, $\epsilon_c = 1$. The slab itself is periodically patterned, that is it consists of alternating regions with permittivity $\epsilon_r = 4$ ($n = 2$) and $\epsilon_r' = 3.24$ ($n' = 1.8$) with a period

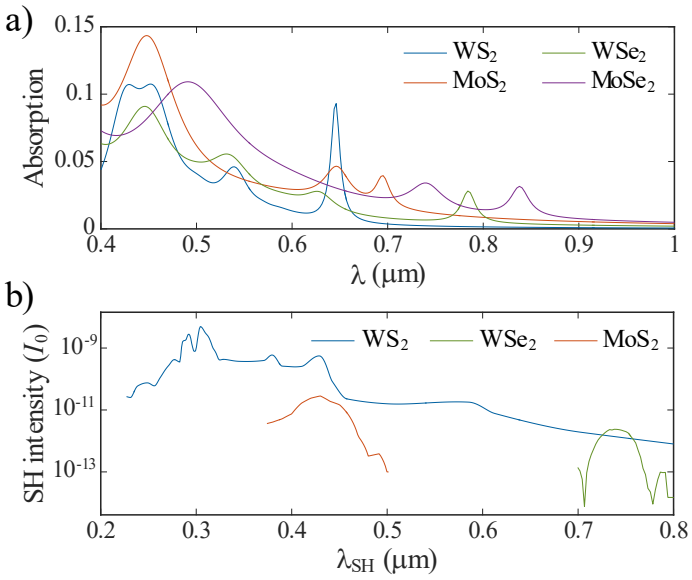

Figure 6.17: a) The FF absorption spectra for ribbons made of WS_2, MoS_2, WSe_2, and $MoSe_2$. b) The SH radiation spectra for the ribbons made of WS_2, MoS_2, and WSe_2.

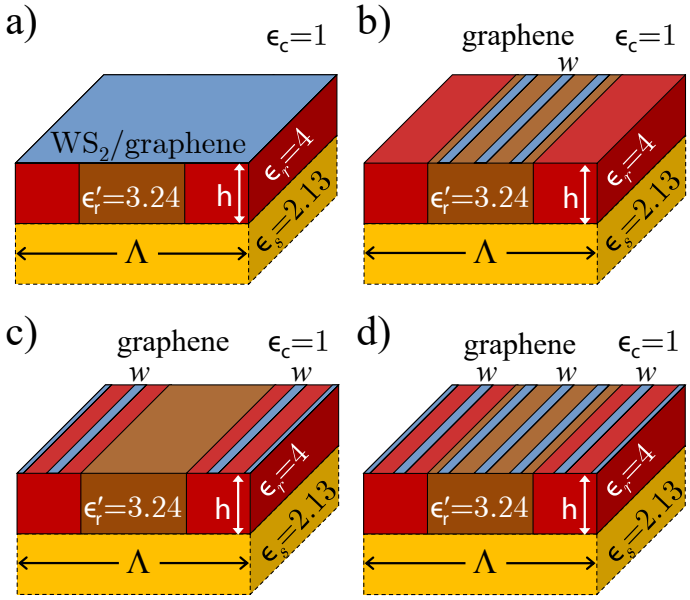

Figure 6.18: Waveguide structures comprising a periodically patterned slab waveguide with permittivities ϵ_r and $\epsilon_r' < \epsilon_r$ covered by air and placed on a dielectric substrate with $\sqrt{\epsilon_s} = 1.46$. On top of the slab waveguide different 2DMs are placed: a) A uniform monolayer of WS_2 or graphene. b), c) Graphene ribbons with width w distributed over the slab material with lower and higher permittivity, respectively. d) Graphene ribbons with width w distributed over both regions of the slab waveguide.

$\Lambda = 400\,\mathrm{nm}$. This particular choice of parameters was inspired by Ref. [42].

For now, consider the unperturbed waveguide consisting of a material with relative permittivity $\bar{\epsilon}_r = (\epsilon_r + \epsilon_r')/2$. A slab waveguide supports optical guided modes and their excitation strongly affects its reflective and transmissive characteristics. A mode of order ν with in-plane propagation constant β_ν can be excited, when the in-plane wavenumber, $k_{||}$, of an incident wave coincides with β_ν. This is not possible for a homogeneous waveguide due to the particular dispersion properties of $k_{||}$ and β_ν, but can be achieved if the waveguide permittivity is periodically modulated, as illustrated in Fig. 6.18(a). This periodic perturbation effectively folds the propagation constant β_ν into the first Brillouin-zone of the k-space of the modes of the periodic waveguide and enables phase matching between $k_{||}$ and β_ν, i.e. the excitation of the mode ν (see also Ref. [43] and Eq. (2.43) in Section 2.6.1). It should be stressed that, albeit less effectively, free-space photons couple to waveguide modes even if the periodic slab waveguide is covered by a thin, optically homogeneous layer, such as a 2DM, because the effective refractive index of the combined structure is periodic in this case, too.

In what follows, the excitation of modes in the TMDC-waveguide structure for a fixed height of $h = 0.18\,\mu\mathrm{m}$ is demonstrated. Subsequently the height of the slab waveguide is optimised to obtain maximal generated SH and the interplay of different resonant mechanisms in the combined waveguide-2D material device is investigated which can lead to an

enhanced nonlinear optical response. Only the monolayer WS_2 is considered as covering 2DM slab in this example, because for this TMDC monolayer the dispersion of the nonlinear conductivity, $\boldsymbol{\sigma}^{(2)}$, is known over the broadest spectral domain. The WS_2 monolayer is oriented such that the arm-chair direction of WS_2 is parallel to the x-axis of the structure. Due to the particular tensorial structure of $\boldsymbol{\sigma}^{(2)}$, this configuration only yields TM-polarised SH for a TM-polarised fundamental field.

In order to demonstrate the excitation of a waveguide mode, consider a TM-polarised, normally incident plane wave in a wavelength range at FF of $\lambda = 0.6\,\mu\mathrm{m}$ to $0.67\,\mu\mathrm{m}$. The corresponding reflection, transmission, and absorption spectra are depicted in Fig. 6.19(a). A spectrally narrow increase of the absorption from less than 0.1 to a maximum of 0.45 is observed at $\lambda = 0.6215\,\mu\mathrm{m}$. Moreover, the transmission and reflection have their minimum and maximum at this wavelength, respectively. This is due to the excitation of the TM_0 waveguide mode, as can be confirmed by the inspection of the electric near-field profile in Fig. 6.19(b): $|E_x|$ has maxima at the top and bottom of the slab waveguide, which also implies a maximum of $|H_y|$ at the centre of the waveguide, as one expects to see for a TM_0 waveguide mode (not shown here). Another local maximum of the absorption can be seen at $\lambda = 0.645\,\mu\mathrm{m}$ and is due to one of the exciton absorption peaks of monolayer WS_2 [c.f. Fig. 6.3(a)].

The enhancement of the fundamental field at the top of the waveguide, where the WS_2-monolayer is located, yields a strongly increased SH radiation with a maximal value of $I_{\mathrm{SH}} = 4 \times 10^{-6} I_0$, as shown in Fig. 6.19(c). Comparison of the radiation spectra for different waveguide heights, $h = 165\,\mathrm{nm}$, $h = 180\,\mathrm{nm}$, and $h = 195\,\mathrm{nm}$, already shows the large sensitivity of the spectral location of waveguide modes to changing height. This is the basis for the parameter study in the remainder of this section. Before that, let us inspect the electric near-field at the SH wavelength, $\lambda_{\mathrm{SH}} = 0.3107\,\mu\mathrm{m}$, presented in Fig. 6.19(d). This profile is markedly different from that of the linear near-field, namely it is spatially more inhomogeneous and has a different distribution of local minima and maxima. Finally, it should be pointed out that the absorption peak at $\lambda = 0.645\,\mu\mathrm{m}$ does no translate to a notable increase of SH radiation, which is similar to the findings reported in Section 6.6.1.

The remainder of this section explores the interplay among different resonant mechanisms in the combined waveguide-2DM structure that lead to enhanced nonlinear optical response and investigate how one can exploit them to increase the generated SH radiation from this photonic device. To this end, TM-polarised incident light in a wavelength range at the FF of $0.2\,\mu\mathrm{m}$ to $0.7\,\mu\mathrm{m}$ is considered and simulations for waveguide height values ranging from $50\,\mathrm{nm}$ to $300\,\mathrm{nm}$, using $N = 25$ orders, are performed. The results of the fundamental frequency calculations are shown in Fig. 6.20 in terms of the reflection spectra, and will be discussed now. The nonlinear part of the device study is summarised in Fig. 6.21

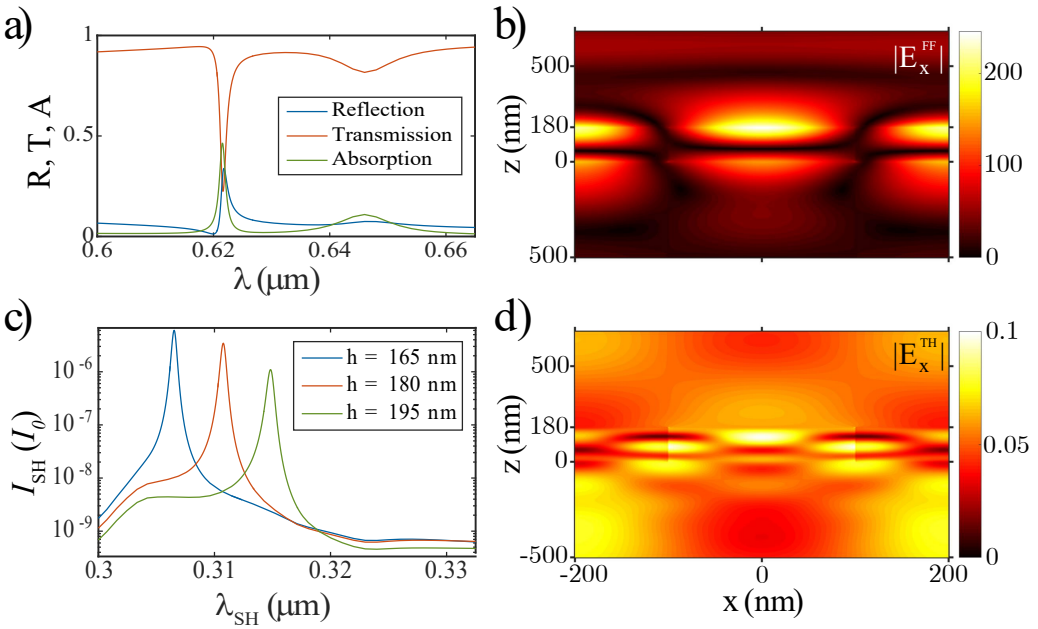

Figure 6.19: a) Linear reflection, absorption, and transmission for waveguide height, $h = 0.18\,\mu\text{m}$, demonstrate the effect of the intrinsic material absorption at $\lambda \approx 0.645\,\mu\text{m}$ and the Fano resonance due to a waveguide resonance at $\lambda = 0.6215\,\mu\text{m}$. b) Electric near-field in and around the waveguide for $\lambda = 0.6215\,\mu\text{m}$ exhibits the spatial profile of the TM_0 mode, with strong field enhancement at waveguide interfaces. c) Nonlinear radiation intensity spectra near the fundamental frequency corresponding to the TM_0 mode, determined for three values of the waveguide height, h. d) Electric near-field at SH wavelength, $\lambda_{\text{SH}} = 0.311\,\mu\text{m}$, for $h = 180\,\text{nm}$.

in terms of its outgoing radiation at the SH and will be investigated subsequently.

The linear optical characteristics of the TMDC-covered waveguide determined for different heights are depicted as a 2D map of reflectivity values. This 2D map of reflection spectra exhibits a smooth dependence on λ and h with minimal and maximal values of $R = 4.9 \times 10^{-6}$ and $R = 0.5428$, respectively, except when one of several mechanisms leads to resonant enhancement of the reflectivity. *i*) The most evident and spectrally broadest features are due to the Fabry-Perot interference mechanism, which yields maximal (minimal) reflectivity if the multiple reflections inside the slab waveguide are in (out-of) phase [45]. These Fabry-Perot resonances appear in the reflectivity map as spectrally broad variations from reflection minima to maxima. *ii)* The second kind of spectral feature is due to the resonant increase of intrinsic optical absorption of WS_2, which occurs at wavelengths at which excitons are generated in the WS_2 monolayer. The strongest of these absorption peaks is at $\lambda = 0.645\,\mu\text{m}$. Their spectral location is determined by the dispersion of $\text{Re}\,(\sigma_s(\omega))$, given in Fig. 6.3(a), and is largely independent of the electromagnetic environment and hence does not depend on h. The intrinsic optical absorption mostly increases the absorption in the combined waveguide-2DM device, but also leads to increasing reflection and decreasing transmission, as was already shown for $h = 180\,\text{nm}$ around $\lambda = 0.65\,\mu\text{m}$

in Fig. 6.19(a). *iii)* The third kind of resonance is due to the excitation of TM_ν waveguide modes and manifests itself as a spectrally narrow and steep variation of the reflectivity of the device. A detailed analysis of the resonant excitation of the TM_0 mode for $h = 180\,\mathrm{nm}$ and $\lambda = 0.6215\,\mu\mathrm{m}$ was already presented in relation to Figs. 6.19. By varying the waveguide height and wavelength, the excitation of the TM_1 and TM_2 modes was found, too.

As the map in Fig. 6.20 suggests, there is a mutual interaction among resonances of the TMDC-covered waveguide, giving rise to several interesting phenomena. First, one can observe the generation of Fano resonances, which generally result from the interference between a discrete state and a broad continuum and are characterised by an asymmetric spectral profile [46–48]. Fano resonances arise via different scenarios in the considered structure, most notably due to the interference of the TM_0 waveguide mode (the discrete state) and the Fabry-Perot resonance (the broad continuum). For example, for a device height of $h = 180\,\mathrm{nm}$ and for increasing wavelengths around $\lambda = 0.62\,\mu\mathrm{m}$, the reflection decreases to a minimum value of $R = 0.01$, then steeply increases to $R = 0.33$, as shown in Fig. 6.19(a). Similar behaviour can be observed when TM_1 and TM_2 waveguide modes are excited. The absorption and transmission spectra exhibit similar features, but the spectral asymmetry, a characteristic feature of Fano resonances, is not as well pronounced in these cases. Hence, only the reflection is shown here. The second phenomenon revealed by Fig. 6.19(a) is the crossing of the TM_0 waveguide mode with the spectrally highest exciton absorption peak of monolayer WS_2. In particular, the two resonances ex-

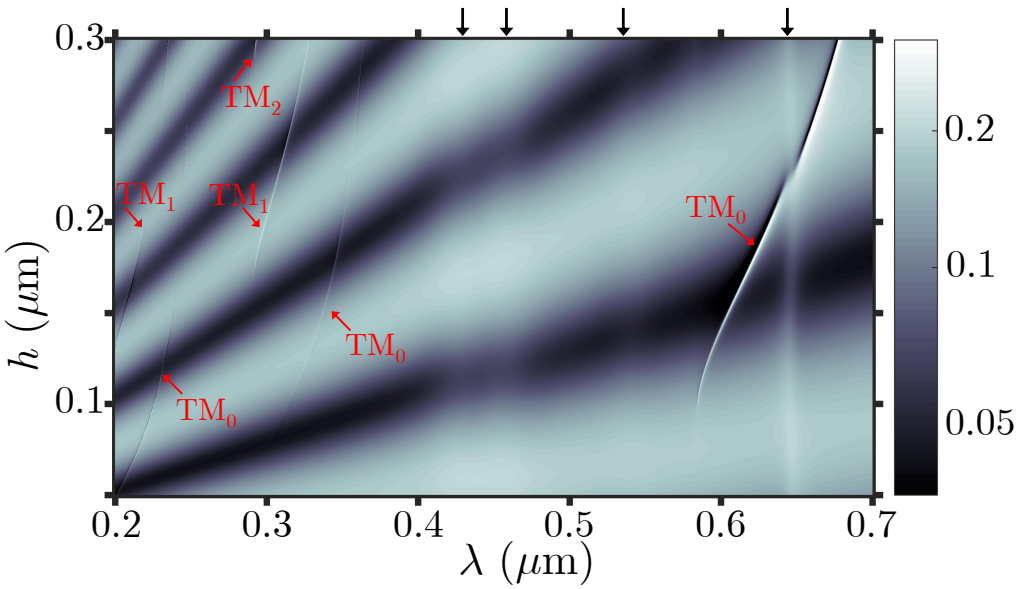

Figure 6.20: Map of the linear reflection spectra for increasing waveguide height, h, exhibits resonances due to the Fabry-Perot interference, the exciton generation (↓), and the excitation of waveguide resonances (↘). Interaction between the TM_0 waveguide mode and the exciton of WS_2 as well as Fano resonances can be observed.

hibit an anti-crossing behaviour at the wavelength of their strongest interaction, which is a well-known phenomenon in photonics and other physical systems [49–51].

Having understood the key features of the linear response of the combined waveguide-2DM optical system, let us now explore its nonlinear optical properties. To this end, consider Fig. 6.21, which depicts the map of the total radiation emitted at the SH wavelengths, λ_{SH}, ranging from $0.1\,\mu\mathrm{m}$ to $0.35\,\mu\mathrm{m}$ and for the same values of the waveguide height as in Fig. 6.20. This intensity varies over almost 6 orders of magnitude, from a minimum of $I_{\mathrm{SH}} = 9\,\mathrm{W\,m^{-2}}$ at $h = 300\,\mathrm{nm}$ and $\lambda_{\mathrm{SH}} = 0.228\,\mu\mathrm{m}$ to a maximum of $I_{\mathrm{SH}} = 6.3 \times 10^{6}\,\mathrm{W\,m^{-2}} \approx 6 \times 10^{-6} I_0$ for $h = 161.5\,\mathrm{nm}$ and $\lambda_{\mathrm{SH}} = 0.306\,\mu\mathrm{m}$. It exhibits a smooth dependence on the system parameters, except for the excitation of certain resonances via mechanisms similar to those examined in the linear case.

In general, the nonlinear radiation is affected by two factors, which will be called *inherited* and *intrinsic* effects. Inherited effects are due to the enhancement via certain mechanisms of the optical field at the FF, at the location of the WS_2 monolayer, which increases the nonlinear source current and consequently the intensity of the generated SH. Intrinsic effects, on the other hand, are resonant effects at the SH wavelength, which can also influence the intensity of radiated waves at the SH.

The most important inherited effects leading to the resonant enhancement of the nonlinear radiation, seen in the map plotted in Fig. 6.21, are as follows: *i)* The inherited Fabry-Perot reflection minima lead to a moderate increase of the SH over broad wavelength ranges. *ii)* The excitation of waveguide modes at the FF leads to particularly strong enhancement of the fundamental field and yields the highest intensity of SH radiation, most notably when

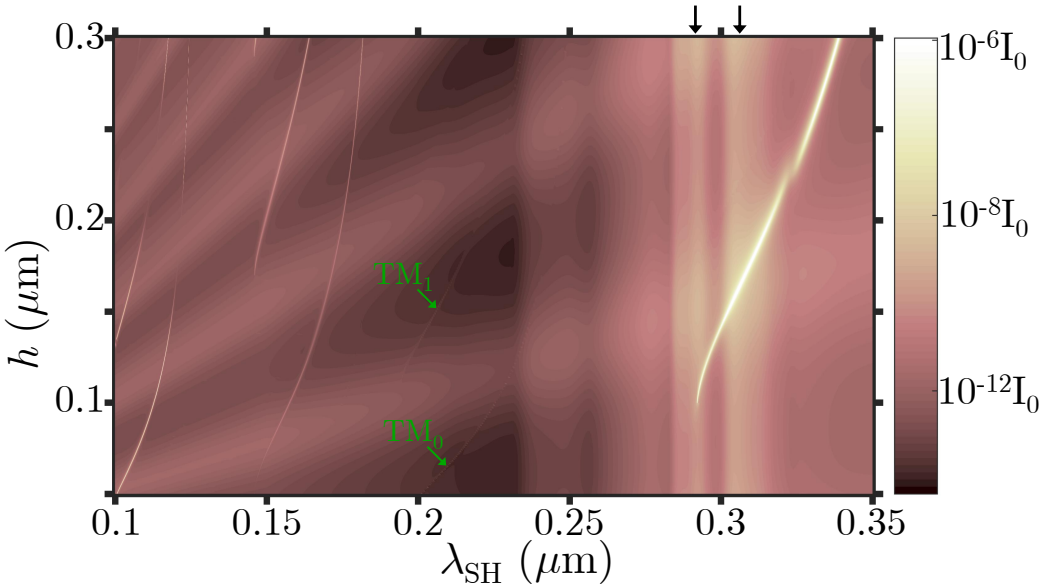

Figure 6.21: Map of nonlinear radiation spectra determined for different waveguide height, h.

the TM_0 mode is excited. In particular, SH radiation with intensity $I_{SH} > 10^{-6} I_0$ is consistently achieved when this mode is excited, except for fundamental wavelengths near the exciton absorption maximum at $\lambda = 0.645\,\mu m$. *iii)* Finally, the interaction of the fundamental TM_0 mode and the WS_2 exciton leads to a reduced enhancement of the fundamental field, as compared to the case of the sole excitation of the TM_0 mode, and results in a decrease of the SH intensity to $I_{SH} = 8.1 \times 10^{-8} I_0$.

Amongst the intrinsic effects, two important mechanisms that lead to enhancement of SH intensity were identified: *i)* The generated electric field at the SH wavelength acts as excitation waves for the TMDC monolayer-waveguide system and resonantly excites waveguide modes existing at the SH, namely the TM_0 and TM_1 modes. This leads to a relatively small increase in the SH intensity. *ii)* The frequency dispersion of the nonlinear optical conductivity of monolayer WS_2 is apparent in the SH spectra: maxima of $\sigma_s^{(2)}$, which are naturally independent of the waveguide height, h, correspond to maxima in the SH radiation spectrum. Moreover, the SHG due to the combined inherited TM_0 mode and the intrinsic maximum of $\sigma_s^{(2)}$ for three values of h is presented in Fig. 6.19(c) and shows that the two effects constructively add to increase the intensity of the SHG.

Note that the reflection, transmission, and absorption spectra, the nonlinear radiation spectra, as well as the resonance wavelengths of the slab waveguide were accurately calculated even for the moderate number of $N = 25$ harmonics, as was ensured with a convergence check for fixed height h and with the rigorous procedures described in Section 6.5.1. A total of 385297 simulations for pairs of (h, λ) were performed, where a higher spectral resolution was used near the resonance wavelengths of the device in order to accurately resolve the spectrally narrow effect of the waveguide resonances.

6.6.3 Nonlinear interaction between waveguide modes and graphene plasmons

The proceeding section combined a 2D material, WS_2, which does not support localised optical modes, with a resonant structure consisting of a periodic slab waveguide, and achieved a strong enhancement of the nonlinear efficiency of the combined device. In this section a similar approach is taken: the resonant periodic slab waveguide is combined with graphene nanostructures, which have already been shown to support localised surface plasmon resonances, in order to achieve a multiresonant, highly nonlinear photonic device. Thus, the structure under consideration is schematically depicted in Fig. 6.18(d). It consists of a periodically patterned slab waveguide with the same optical parameters as the one in the preceding section, which now is covered by graphene ribbons with width $w = 230\,nm$. Let us call the three graphene ribbons centred on top of the material with permittivity ϵ_r' and ϵ_r the inner and outer ribbons, respectively, which is natural given the definition of the unit cell in Fig. 6.18(d). The centre-to-centre distance of the inner ribbons (and outer ribbons) is

$3w = 0.69\,\mu\mathrm{m}$ and the centre-to-centre distance between a inner ribbon to a neighbouring outer ribbon is $1.37\,\mu\mathrm{m}$. In this case, the height of the slab waveguide is $h = 1.5\,\mu\mathrm{m}$ and the period is $\Lambda = 5.5\,\mu\mathrm{m}$.

The small feature size of the graphene ribbons, $w = 0.0418\Lambda$, is required to excite graphene plasmons at moderately small wavelengths, at which waveguide modes exist, too. This is not a conceptual drawback of this particular structure, but it is computationally costly to accurately resolve graphene ribbons with very small width. Thus, $N = 251$ harmonics were used throughout this computational analysis and an additional conductivity of $\sigma_{s,\mathrm{add}}(10^{-3})$ was introduced.

To fully understand the linear and nonlinear optical response of the graphene-waveguide structure, let us first investigate two less complex, complementary structures where the graphene ribbons are located on either the inner or the outer part of the waveguide, as per Figs. 6.18(b) and 6.18(c), respectively, and also the waveguide covered with an unstructured, uniform graphene sheet, as per Fig. 6.18(a).

The linear absorption spectra for these three variations of the device, determined for values of the fundamental wavelength ranging from $1\,\mu\mathrm{m}$ to $20\,\mu\mathrm{m}$, are depicted in Fig. 6.22(a), where normal incidence and TM-polarisation is assumed. The absorption of the waveguide with the covering graphene sheet follows a monotonously increasing trend, upon which alternating, broad local minima and maxima are superimposed. These maxima are due to the Fabry-Perot interference and primarily reveal themselves as maxima and minima of the device reflectivity, also shown in Fig. 6.22(a). The absorption spectra for the waveguide with graphene ribbons on top of the inner and outer parts of the waveguide exhibit relatively broad spectral peaks due to the excitation of surface plasmons in the graphene ribbons. Their excitation wavelength chiefly depends on the width of the ribbon and the permittivity of the underlying dielectric: the lower refractive index material ($n' = 1.8$) underneath the inner ribbons leads to excitation of plasmons at smaller wavelength than the wavelength corresponding to the plasmons in the outer ribbons, placed on top of waveguide sections with higher refractive index ($n = 4$). In all three devices, the absorption exhibits additional, spectrally narrow peaks with up to 30% absorption due to the excitation of optical modes in the slab waveguide grating.

The nonlinear radiation spectra in Fig. 6.22(b) complement these findings. Thus, the intensity of the TH radiation generated by the unstructured graphene sheet increases with wavelength and one can observe the effect of Fabry-Perot resonances overlaid on it. Moreover, the local field enhancement due to the excitation of localised surface plasmons in the inner and outer graphene ribbons leads to increased THG at lower and higher wavelengths, respectively. The inherited waveguide modes maximise the amount of generated TH radiation, which reaches values of up to $I_{\mathrm{TH}} = 2 \times 10^{-3} I_0$.

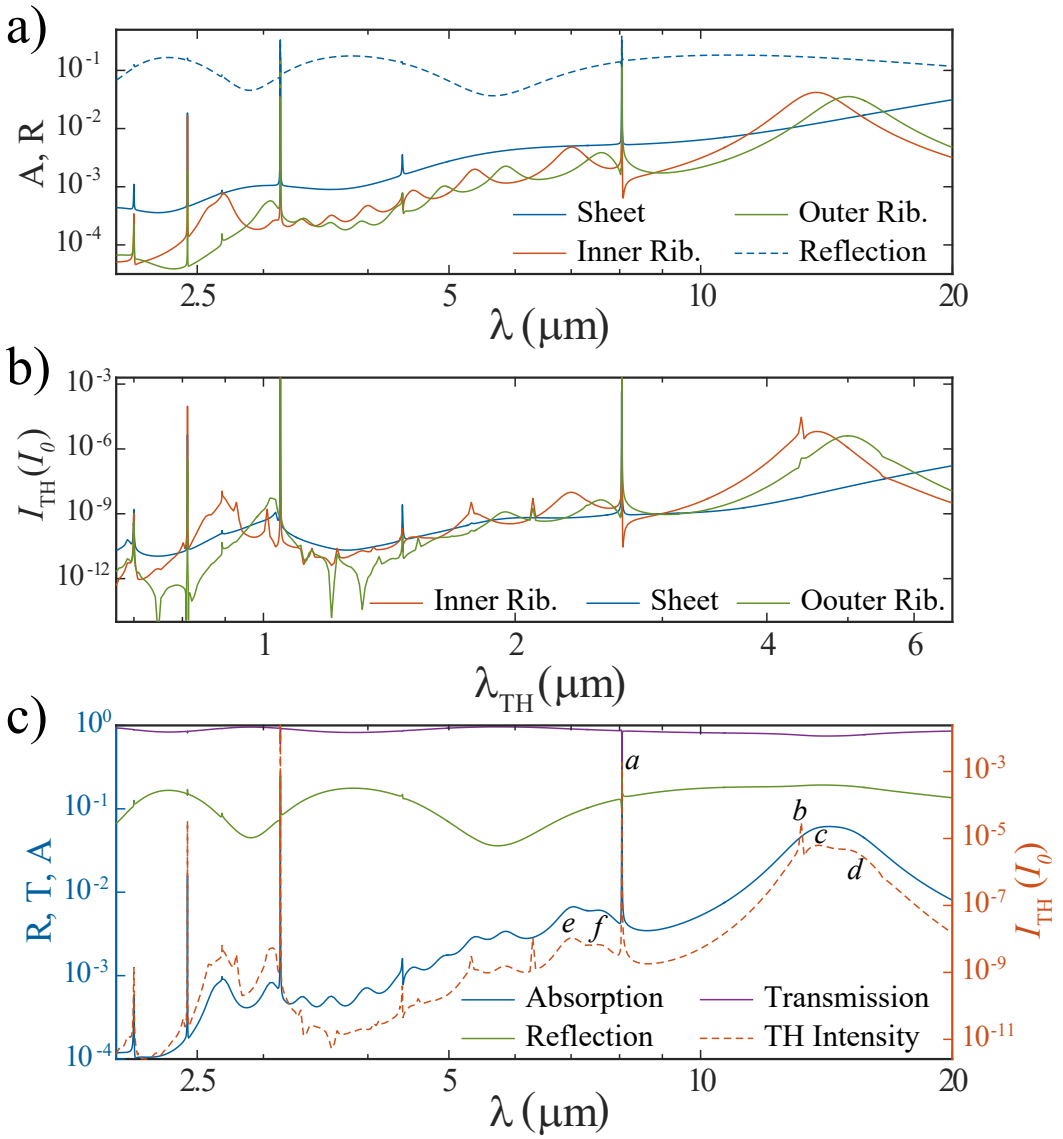

Figure 6.22: a) Absorption spectra (solid lines) of the homogeneous graphene sheet and the graphene ribbons on top of inner and outer parts of the slab waveguide, c.f. Figs. 6.18(a) through 6.18(c), and reflection spectrum (dashed line) of the homogeneous graphene sheet. b) The nonlinear radiation spectrum of the same structures as in a). c) Linear spectra (reflection, transmission, and absorption) and TH spectrum for the graphene ribbons placed on top of both parts of the waveguide, as per Fig. 6.18(d).

The interplay of these resonant effects in the combined graphene-waveguide structure with ribbons distributed over the whole area of the waveguide is revealed by the results presented in Fig. 6.22(c). Note that the abscissa of this figure gives the values of the incoming wavelength, λ, for the linear reflection, transmission, and absorption spectra and the TH wavelength, $\lambda_{\mathrm{TH}} = \lambda/3$, for the intensity of THG. The absorption at the FF displays an increasing trend in the range of $2\,\mu\mathrm{m}$ to $20\,\mu\mathrm{m}$; however, there are several spectrally broad and narrow absorption peaks, which are the manifestation of different phenomena. Thus,

the narrow resonances are due to the excitation of the TM_0 mode of the slab waveguide, e.g. those at $\lambda = 8.05\,\mu\mathrm{m}$, $\lambda = 4.4\,\mu\mathrm{m}$, and $\lambda = 3.14\,\mu\mathrm{m}$. The electric field, E_x^{FF}, at the FF, $\lambda = 8.05\,\mu\mathrm{m}$, shows a strong enhancement at the top of the waveguide region, as per Fig. 6.23(a, top). Similarly to the TM_0 mode shown in Fig. 6.19(b), this field profile has two maxima over the x-extent of one unit cell of the grating and two maxima along the z-extent of the waveguide, where the maximum near the top of the waveguide is much larger than the one at the bottom. This strong fundamental field enhancement increases the absorption to 33% and leads to a strong nonlinear source current, which in turn generates a strong electric field at the TH, the near-field of which is depicted in Fig. 6.23(a, bottom). This TH field is mostly localised around the graphene ribbons and has an evanescent nature. The total intensity of TH radiated into the cover and substrate amounts to $I_{\mathrm{TH}} = 1.3 \times 10^{-3} I_0$.

The excitation of the TM_0 mode at $\lambda = 3.14\,\mu\mathrm{m}$ is equally interesting. It appears both as sharp local maximum in the linear absorption spectrum at $\lambda = 3.14\,\mu\mathrm{m}$ and as an increase of TH intensity at the TH wavelength $\lambda_{\mathrm{TH}} = 3.14\,\mu\mathrm{m}$ in Fig. 6.22(c). The nonlinear electric near-field profile in Fig. 6.23(b, bottom part) confirms the excitation of this intrinsic nonlinear TM_0 mode for $\lambda_{\mathrm{TH}} = 3.14\,\mu\mathrm{m}$.

Another reason for increased absorption is the excitation of localised surface plasmons on the graphene ribbons. In order to illustrate this, let us consider the two absorption peaks with largest wavelengths, around $\lambda \approx 14\,\mu\mathrm{m}$ and $\lambda \approx 7.3\,\mu\mathrm{m}$. The absorption maximum with the largest wavelength appears in the linear spectrum as a very broad resonance around $\lambda \approx 14\,\mu\mathrm{m}$, but the electric near-field profiles in the two top panels of Figs. 6.23(c) and 6.23(d) reveal that actually the surface plasmons are excited on the inner and outer graphene ribbons at slightly different wavelengths, $\lambda = 13.82\,\mu\mathrm{m}$ and $\lambda = 15.17\,\mu\mathrm{m}$, respectively, which is explained by the difference in the electromagnetic environment probed by the corresponding plasmons. The excitation of these surface plasmons can also be seen in Fig. 6.22(c), as the local maxima labelled by "*c*" and "*d*". Moreover, they can be viewed directly in the nonlinear far-field radiation spectrum, too, as spectrally separated peaks. As a consequence of the excitation of localised surface plasmons, the TH near-field shows its highest values near the inner and outer ribbons at the TH wavelengths, $\lambda_{\mathrm{TH}} = 13.82\,\mu\mathrm{m}/3$ and $\lambda_{\mathrm{TH}} = 15.17\,\mu\mathrm{m}/3$, as illustrated in the bottom panels of Figs. 6.23(c) and 6.23(d), respectively.

Similar physics describe the lower-order plasmon corresponding to $\lambda \approx 7.3\,\mu\mathrm{m}$. Thus, by inspecting the profiles of the fundamental near-field in the top panels of Figs. 6.23(e) and 6.23(f), one can see that, correspondingly, a plasmon with three peaks in the field profile is excited on the inner ribbons at $\lambda = 7.1\,\mu\mathrm{m}$ whereas this same type of plasmon is excited on the outer ribbons at $\lambda = 7.6\,\mu\mathrm{m}$. This near-field pattern is in accordance to the surface plasmon field profiles that were found in Section 6.5.2, Fig. 6.16(b). Moreover, no optical

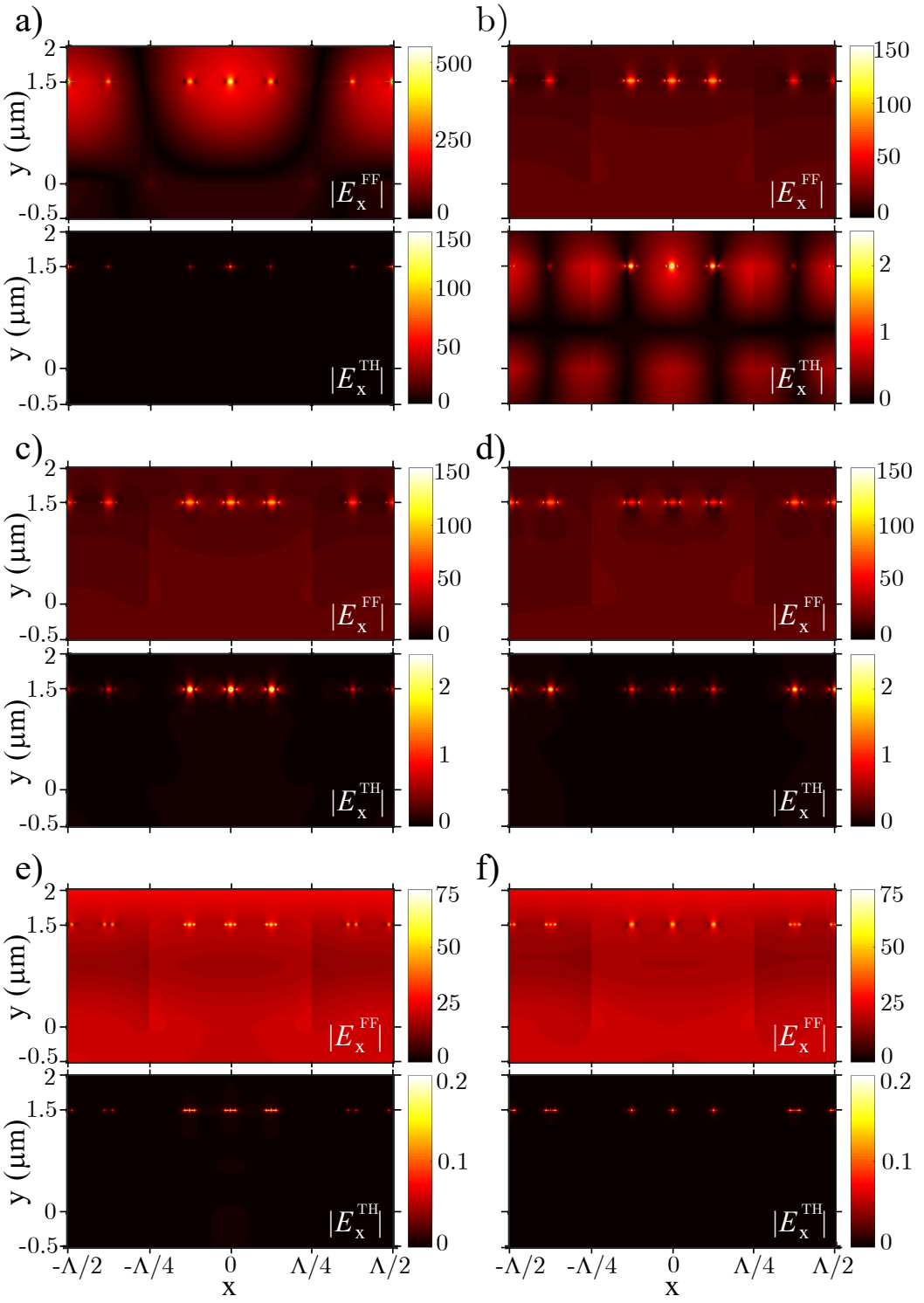

Figure 6.23: Dominant component of the electric field E_x at the FF (top parts) and the TH (bottom parts) for selected fundamental wavelengths λ and TH wavelengths $\lambda_{\mathrm{TH}} = \lambda/3$: a) Field enhancement due to excitation of the TM_0 waveguide mode for $\lambda = 8.05\,\mu\mathrm{m}$. b) Near-field profile of the TM_0 mode at the TH wavelength $\lambda_{\mathrm{TH}} = 4.4\,\mu\mathrm{m}$. c), d) Plasmonic field enhancement in the inner and outer ribbons for $\lambda = 13.82\,\mu\mathrm{m}$ and $\lambda = 15.17\,\mu\mathrm{m}$, respectively. e), f) Excitation of surface plasmons in graphene ribbons located on top of waveguide sections with ϵ_r' (inner ribbons, $\lambda = 7.1\,\mu\mathrm{m}$) and ϵ_r (outer ribbons, $\lambda = 7.6\,\mu\mathrm{m}$), respectively. The labels of the panels correspond to the labels of the peaks in Fig. 6.22(c).

coupling between the plasmonic fields near adjacent ribbons could be observed. The spectra of the TH radiation at the wavelengths $\lambda_{\mathrm{TH}} = 7.1\,\mu\mathrm{m}/3$ and $\lambda_{\mathrm{TH}} = 7.6\,\mu\mathrm{m}/3$ also exhibits two local maxima, due to the enhancement of the fundamental field. The number of electric field maxima near the inner and outer graphene ribbons in the bottom of Figs. 6.23(e) and 6.23(f), respectively, agrees with the 5 field maxima in Fig. 6.16(d).

The two physical mechanisms that lead to the resonant enhancement of the optical response of the graphene-waveguide structure can partly be observed in the reflection and transmission spectra at the fundamental wavelength, which mainly exhibit a variation between minima and maxima due to the Fabry-Perot interference. This pattern is complemented by spectrally very narrow regions of increased reflection and decreased transmission due to the excitation of waveguide modes. The influence of surface plasmons on the reflection and transmission is only apparent at the largest excitation wavelengths around $\lambda \approx 14\,\mu\mathrm{m}$, where reflection and transmission are notably increased and decreased, respectively.

6.6.4 Tuning the nonlinear interaction of waveguide modes and graphene plasmons

An effective way of tuning the characteristics of the light radiated by our nonlinear diffraction grating is by varying the angle of incidence of the incoming light. Not only will this showcase the effectiveness of the proposed numerical method in the oblique-incidence configuration, but it will also demonstrate the influence of the angle of incidence on the optical response of the grating. For example, the interaction of the spectrally narrow TM_0 waveguide mode at $\lambda = 8.05\,\mu\mathrm{m}$ with the two broad continua corresponding to the excitation of graphene plasmons at $\lambda = 7.1\,\mu\mathrm{m}$ and $\lambda = 7.6\,\mu\mathrm{m}$ provides a convenient optical setting for studying a tunable Fano resonance resulting from the interaction between a discrete state and multiple continua, a phenomenon that has recently been explored in a different plasmonic structure [52].

In order to investigate the interaction between a waveguide mode and the double absorption peak, 105747 simulations with $N = 251$ harmonics have been performed for increasing angle of incidence, θ, ranging from 0° to 20° (and constant azimuthal angle $\varphi = 0^\circ$) in the fundamental wavelength range of the second highest plasmon peaks, namely from $6\,\mu\mathrm{m}$ to $9\,\mu\mathrm{m}$. Higher spectral resolution was employed near the waveguide resonance wavelengths of the slab waveguide.

The resulting absorption map is shown in Fig. 6.24. The spectrum for $\theta = 0^\circ$ is the same as in Fig. 6.22(c) and shows two broad maxima of $A \approx 6 \times 10^{-3}$, due to the excitation of surface plasmons, and a sharp maximum of $A = 0.33$ due to the excitation of the TM_0 waveguide mode at $\lambda^{\mathrm{TM}_0} = 8.05\,\mu\mathrm{m}$. Increasing the incident angle, θ, leaves the spectral location of the plasmon excitation unchanged; however, the spectral location

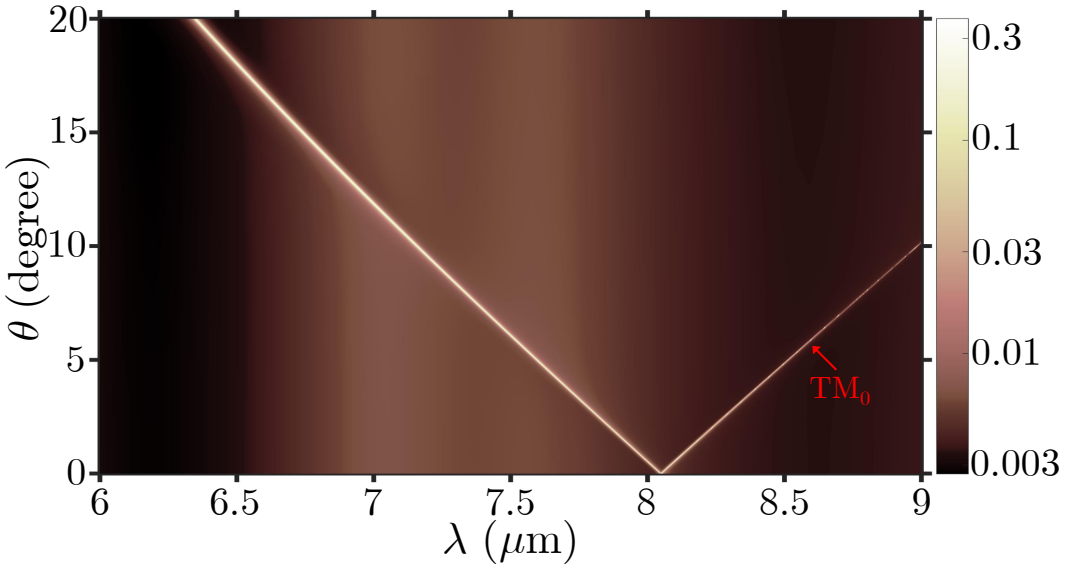

Figure 6.24: Map of absorption spectra vs. angles of incidence, θ, determined in the wavelength range of the second-order surface plasmon of graphene ribbons.

of the waveguide mode resonance varies. For $\theta > 0$, the TM_0 mode is excited at two wavelengths, $\lambda^{TM_0^l}(\theta) < \lambda^{TM_0}$ and $\lambda^{TM_0^r}(\theta) > \lambda^{TM_0}$. Their separation from λ^{TM_0} increases with θ and a more effective excitation on the left branch of the TM_0 mode can be seen as compared to the right branch. For $\theta = 4.96°$ and $\theta = 10.71°$ the wavelengths of the mode, $\lambda^{TM_0^l}(10.71°) = 7.1\,\mu m$ and $\lambda^{TM_0^l}(4.96°) = 7.6\,\mu m$, coincide with the central wavelength of the plasmon absorption peaks of the inner and outer ribbons, yielding absorption of $A = 0.331$ and $A = 0.238$, respectively.

Along the TM_0 band, i.e. the path of excitation of the TM_0 mode in the (λ, θ)-space, strong enhancement of conversion efficiency due to inherited effect of this mode can be observed in Fig. 6.25. The simultaneous excitation of waveguide modes and graphene plasmons yields strong THG, with intensities of $I_{TH} = 5.85 \times 10^{-5} I_0$ at $\lambda_{TH} = 2.37\,\mu m$ and $\theta = 10.71°$ and $I_{TH} = 3.89 \times 10^{-4} I_0$ at $\lambda_{TH} = 2.53\,\mu m$ and $\theta = 4.96°$. However, the strongest TH intensity of $I_{TH} = 0.0537 I_0$ is generated at $\lambda_{TH} = 2.42\,\mu m$ and $\theta = 9°$, namely where the intrinsic TM_1 band crosses the inherited TM_0 band. The increase of the conversion efficiency due to sole excitation of the intrinsic nonlinear modes is notable, but orders of magnitude lower than in the case of inherited effects, e.g. the same intrinsic TM_1 mode away from the simultaneous resonance, at $\lambda_{TH} = 2.16\,\mu m$ and $\theta = 20°$, yields TH radiation of intensity $I_{TH} = 7.13 \times 10^{-9} I_0$.

Inspection of the bands of the intrinsic modes reveals and additional interesting feature. Specifically, intrinsic modes of the same order show anti-crossing behaviour, e.g. TM_0 at $\lambda_{TH} = 2.75\,\mu m$ and $\theta = 14.42°$ or TM_1 at $\lambda_{TH} = 2.33\,\mu m$ and $\theta = 12.25°$, whereas

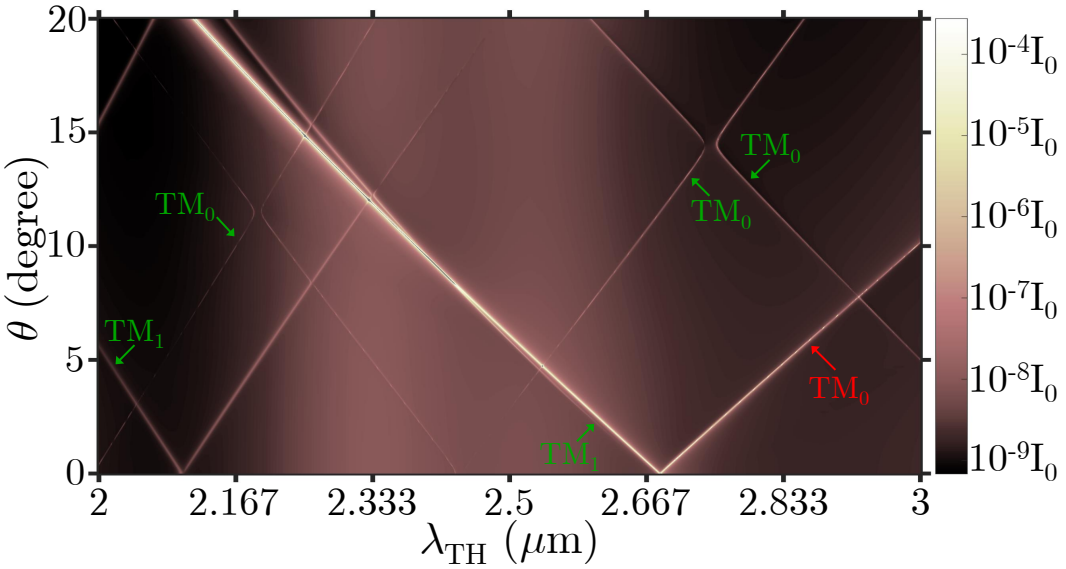

Figure 6.25: The map of nonlinear radiation spectra shows the inherited (red labels) and intrinsic, nonlinear (green labels) modes of the slab waveguide. The left and right branches of intrinsic modes of the same order exhibit anti-crossing mode interaction, which leads to the formation of spectral band-gaps.

bands of different intrinsic modes pass through each other, i.e. TM_0 and TM_1 at $\lambda_{\mathrm{TH}} = 2.26\,\mu\mathrm{m}$ and $\theta = 8.375°$. This crossing of modes of same order occurs when the transverse component of the incident k vector, $k_{||}$, reaches the edge of the first Brillouin zone.

6.7 Conclusions

This chapter derived an accurate formulation of the RCWA method to describe linear and nonlinear interaction of light with periodically patterned 2D materials. This numerical formalism does not depend on the height of a 2D material and is applicable to different 2D materials, as long as their linear and nonlinear surface conductivities are known. A crucial ingredient for the accuracy and convergence of this numerical method was the introduction of small, additional conductivity, which allows the correct solution of the Fourier factorisation problem and hence the correct numerical handling of 2D materials. A small value of the added conductivity $\sigma_{s,\mathrm{add}}(\eta)$ yields accurate results when convergence with respect to the number of harmonics N has been achieved. However, this is computationally costly, as high N are required. This can be seen as an opportunity to substantially speed up the calculation by using a slightly higher additional conductivity, at the cost of a small introduced error. The presented method also allows the prediction of the nonlinear optical response of periodic 2D materials and it was found that correct nonlinear results can only be obtained with the accurate near-field formulation of RCWA from Chapter 4.

The numerical method was validated by comparing the results to a reference method,

and subsequently applied to investigate the characteristics of gratings comprising different 2D materials. We found that 2D materials have different linear and nonlinear properties. Graphene supports surface plasmons and TH radiation due to its conductive nature and symmetry properties, respectively. The surface plasmons were found to increase linear absorption and to enhance THG in graphene, which reveals the potential for tuneable THG in graphene. The TMDC monolayer materials, a different kind of 2D materials, are dielectric and centrosymmetric, and they were found to not support intrinsic resonances. However, by coupling a nonresonant TMDC monolayer with a 3D resonant structure, namely a periodically patterned waveguide, a way to increase the nonlinear conversion efficiency of the TMDC monolayer at selected wavelengths was shown. By combining the already resonant graphene with a similar waveguide structure, the interplay of their resonances, namely the surface plasmons and the waveguide resonances, was investigated, and the increase of conversion efficiency due to the excitation of Fano resonances was shown.

In this work, only the 2D materials in the grating have been considered to possess an optical nonlinearity. A numerical method which allows the calculations of nonlinear effects in the remaining bulk parts of a grating structure will be presented in the following.

Bibliography

[1] K. S. Novoselov, A. K. Geim, S. Morozov, D Jiang, Y Zhang, S. Dubonos, I. Grigorieva, and A. Firsov, "Electric field effect in atomically thin carbon films," Science **306**, 666–669 (2004).

[2] K. Novoselov, A. K. Geim, S. Morozov, D Jiang, M. Katsnelson, I. Grigorieva, S. Dubonos, and A. Firsov, "Two-dimensional gas of massless Dirac fermions in graphene," Nature **438**, 197–200 (2005).

[3] K. S. Novoselov, D. Jiang, F. Schedin, T. J. Booth, V. V. Khotkevich, S. V. Morozov, and A. K. Geim, "Two-dimensional atomic crystals," Proc. Natl. Acad. Sci. USA **102**, 10451–10453 (2005).

[4] W. Gao, J. Shu, C. Qiu, and Q. Xu, "Excitation of plasmonic waves in graphene by guided-mode resonances," ACS Nano **6**, 7806–7813 (2012).

[5] J. Gosciniak and D. T. Tan, "Theoretical investigation of graphene-based photonic modulators," Sci. Rep. **3**, 1897 (2013).

[6] A. Y. Nikitin, F. Guinea, F. J. Garcia-Vidal, and L. Martin-Moreno, "Surface plasmon enhanced absorption and suppressed transmission in periodic arrays of graphene ribbons," Phys. Rev. B **85**, 081405 (2012).

[7] I. Llatser, C. Kremers, A. Cabellos-Aparicio, J. M. Jornet, E. Alarcon, and D. N. Chigrin, "Graphene-based nano-patch antenna for terahertz radiation," Phot. Nano. Fund. Appl. **10**, 353–358 (2012).

[8] A. Khavasi, "Fast convergent fourier modal method for the analysis of periodic arrays of graphene ribbons," Opt. Lett. **38**, 3009–3012 (2013).

[9] V. Nayyeri, M. Soleimani, and O. M. Ramahi, "Modeling Graphene in the Finite-Difference Time-Domain Method using a Surface Boundary Condition," IEEE Trans. Antennas Propag. **61**, 4176–4182 (2013).

[10] P.-Y. Chen and A. Alu, "Atomically thin surface cloak using graphene monolayers," ACS Nano **5**, 5855–5863 (2011).

[11] S. A. H. Nekuee, A. Khavasi, and M. Akbari, "Fourier modal method formulation for fast analysis of two-dimensional periodic arrays of graphene," J. Opt. Soc. Am. B **31**, 987–993 (2014).

[12] B. Wunsch, T. Stauber, F. Sols, and F. Guinea, "Dynamical polarization of graphene at finite doping," New J. Phys. **8**, 318 (2006).

[13] T. C. Berkelbach, M. S. Hybertsen, and D. R. Reichman, "Theory of neutral and charged excitons in monolayer transition metal dichalcogenides," Phys. Rev. B **88**, 045318 (2013).

[14] J. L. Cheng, N Vermeulen, and J. E. Sipe, "Third order optical nonlinearity of graphene," New J. Phys. **16**, 053014 (2014).

[15] C. Janisch, Y. Wang, D. Ma, N. Mehta, A. L. Elas, N. Perea-Lpez, M. Terrones, V. Crespi, and Z. Liu, "Extraordinary Second Harmonic Generation in Tungsten Disulfide Monolayers," Sci. Rep. **4**, 5530 (2014).

[16] Y. Li, A. Chernikov, X. Zhang, A. Rigosi, H. M. Hill, A. M. van der Zande, D. A. Chenet, E.-M. Shih, J. Hone, and T. F. Heinz, "Measurement of the optical dielectric function of monolayer transition-metal dichalcogenides: MoS_2, $MoSe_2$, WS_2, and WSe_2," Phys. Rev. B **90**, 205422 (2014).

[17] B. Mukherjee, F. Tseng, D. Gunlycke, K. K. Amara, G. Eda, and E. Simsek, "Complex electrical permittivity of the monolayer molybdenum disulfide (MoS_2) in near UV and visible," Opt. Mater. Express **5**, 447–455 (2015).

[18] S.-Y. Hong, J. I. Dadap, N. Petrone, P.-C. Yeh, J. Hone, and R. M. Osgood Jr, "Optical third-harmonic generation in graphene," Phys. Rev. X **3**, 021014 (2013).

[19] F. H. Koppens, D. E. Chang, and F. J. Garcia de Abajo, "Graphene plasmonics: a platform for strong light–matter interactions," Nano Lett. **11**, 3370–3377 (2011).

[20] Y. Francescato, V. Giannini, and S. A. Maier, "Strongly confined gap plasmon modes in graphene sandwiches and graphene-on-silicon," New J. Phys. **15**, 063020 (2013).

[21] J. Ribeiro-Soares, R. M. Almeida, E. B. Barros, P. T. Araujo, M. S. Dresselhaus, L. G. Can çado, and A. Jorio, "Group theory analysis of phonons in two-dimensional transition metal dichalcogenides," Phys. Rev. B **90**, 115438 (2014).

[22] Y. Shen, *The principles of nonlinear optics*, Wiley classics library (Wiley-Interscience, 2003).

[23] R. W. Boyd, *Nonlinear optics* (Academic Press, 2008).

[24] L. M. Malard, T. V. Alencar, A. P. M. Barboza, K. F. Mak, and A. M. de Paula, "Observation of intense second harmonic generation from MoS_2 atomic crystals," Phys. Rev. B **87**, 201401 (2013).

[25] Y. Li, Y. Rao, K. F. Mak, Y. You, S. Wang, C. R. Dean, and T. F. Heinz, "Probing Symmetry Properties of Few-Layer MoS_2 and h-BN by Optical Second-Harmonic Generation," Nano Lett. **13**, 3329–3333 (2013).

[26] K. L. Seyler, J. R. Schaibley, P. Gong, P. Rivera, A. M. Jones, S. Wu, J. Yan, D. G. Mandrus, W. Yao, and X. Xu, "Electrical control of second-harmonic generation in a WSe_2 monolayer transistor," Nat. Nanotechnol. **10**, 407–411 (2015).

[27] S. Bergfeld and W. Daum, "Second-Harmonic Generation in GaAs: Experiment versus Theoretical Predictions of $\chi^{(2)}_{xyz}$," Phys. Rev. Lett. **90**, 036801 (2003).

[28] D. Krause, C. W. Teplin, and C. T. Rogers, "Optical surface second harmonic measurements of isotropic thin-film metals: Gold, silver, copper, aluminum, and tantalum," J. Appl. Phys. **96**, 3626–3634 (2004).

[29] W. Nakagawa, R.-C. Tyan, and Y. Fainman, "Analysis of enhanced second-harmonic generation in periodic nanostructures using modified rigorous coupled-wave analysis in the undepleted-pump approximation," J. Opt. Soc. Am. A. **19**, 1919–1928 (2002).

[30] B. Bai and J. Turunen, "Fourier modal method for the analysis of second-harmonic generation in two-dimensionally periodic structures containing anisotropic materials," J. Opt. Soc. Am. B. **24**, 1105–1112 (2007).

[31] T. Paul, C. Rockstuhl, and F. Lederer, "A numerical approach for analyzing higher harmonic generation in multilayer nanostructures," J. Opt. Soc. Am. B. **27**, 1118–1130 (2010).

[32] J. D. Jackson, *Classical Electrodynamics*, 3rd ed. (John Wiley, 1999).

[33] P. Lalanne and G. M. Morris, "Highly improved convergence of the coupled-wave method for tm polarization," J. Opt. Soc. Am. A. **13**, 779–784 (1996).

[34] L. Li, "New formulation of the fourier modal method for crossed surface-relief gratings," J. Opt. Soc. Am. A. **14**, 2758–2767 (1997).

[35] L. Li, "Fourier modal method for crossed anisotropic gratings with arbitrary permittivity and permeability tensors," J. Opt. Soc. Am. A. **5**, 345 (2003).

[36] 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**, 2880–2890 (2007).

[37] M. Born and E. Wolf, *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light* (Cambridge University Press, 1999) Chap. 11.5.

[38] E. Popov, M. Neviere, B. Gralak, and G. Tayeb, "Staircase approximation validity for arbitrary-shaped gratings," J. Opt. Soc. Am. A. **19**, 33–42 (2002).

[39] P. Lalanne and M. P. Jurek, "Computation of the near-field pattern with the coupled-wave method for transverse magnetic polarization," J. Mod. Opt. **45**, 1357–1374 (1998).

[40] M. Weismann, D. F. G. Gallagher, and N. C. Panoiu, "Accurate near-field evaluation in the rigorous coupled-wave analysis," J. Opt. **17**, 125612 (2015).

[41] MATLAB, *version 8.6 (R2015b)* (The MathWorks Inc., Natick, Massachusetts, USA, 2015).

[42] S. Peng and G. M. Morris, "Resonant scattering from two-dimensional gratings," J. Opt. Soc. Am. A. **13**, 993–1005 (1996).

[43] N. C. Panoiu and R. M. Osgood, "Enhanced optical absorption for photovoltaics via excitation of waveguide and plasmon-polariton modes," Opt. Lett. **32**, 2825–2827 (2007).

[44] M. Weismann, D. F. G. Gallagher, and N. C. Panoiu, "Nonlinear generalized source method for modeling second-harmonic generation in diffraction gratings," J. Opt. Soc. Am. B. **32**, 523–533 (2015).

[45] G. Hernández, *Fabry-perot interferometers*, Vol. 3 (Cambridge University Press, 1988).

[46] U. Fano, "Effects of configuration interaction on intensities and phase shifts," Phys. Rev. **124**, 1866–1878 (1961).

[47] A. E. Miroshnichenko, S. Flach, and Y. S. Kivshar, "Fano resonances in nanoscale structures," Rev. Mod. Phys. **82**, 2257 (2010).

[48] B. Luk'yanchuk, N. I. Zheludev, S. A. Maier, N. J. Halas, P. Nordlander, H. Giessen, and C. T. Chong, "The Fano resonance in plasmonic nanostructures and metamaterials," Nat. Mater. **9**, 707–715 (2010).

[49] C. Weisbuch, M. Nishioka, A Ishikawa, and Y Arakawa, "Observation of the coupled exciton-photon mode splitting in a semiconductor quantum microcavity," Phys. Rev. Lett. **69**, 3314 (1992).

[50] T. Yoshie, A. Scherer, J Hendrickson, G Khitrova, H. Gibbs, G Rupper, C Ell, O. Shchekin, and D. Deppe, "Vacuum Rabi splitting with a single quantum dot in a photonic crystal nanocavity," Nature **432**, 200–203 (2004).

[51] L. Novotny, "Strong coupling, energy splitting, and level crossings: A classical perspective," Am. J. Phys. **78**, 1199–1202 (2010).

[52] E. J. Osley, C. G. Biris, P. G. Thompson, R. R. F. Jahromi, P. A. Warburton, and N. C. Panoiu, "Fano Resonance Resulting from a Tunable Interaction between Molecular Vibrational Modes and a Double Continuum of a Plasmonic Metamolecule," Phys. Rev. Lett. **110**, 087402 (2013).

Chapter 7

Numerical modelling of optical nonlinearities in periodic structures using the generalised source method

7.1 Introduction

This chapter introduces a versatile numerical method for modelling the diffraction of light in periodically patterned photonic structures containing different types of nonlinear optical materials. The presented approach extends the linear generalised source method [1–3] to nonlinear optical interactions by incorporating the contribution of nonlinear polarization sources to the diffracted field in the algorithm. The linear GSM is chosen as a starting platform here because it has proven to be an efficient alternative to the RCWA for the analysis of one- and two-dimensional (1D, 2D), corrugated, periodic optical structures. Its optimal runtime is of order $N_0 \log(N_0)$, where N_0 is the total number of diffraction orders, as compared to N_0^3 for the conventional RCWA. The key ingredients needed to achieve this remarkable computational efficiency is the use of a matrix-multiplication based iterative method to solve an underlying linear system of coupled integral equations and a fast Toeplitz matrix-vector multiplication performed by means of a fast Fourier transform. In addition, its implicit, Green's function type formulation of Maxwell equations allows for a natural incorporation of nonlinear source terms, which is the main challenge in computing the nonlinear response of optical systems.

The remainder of this chapter is organized as follows: in Section 7.2 the general idea of the nonlinear GSM is presented and the mathematical formulation of SHG in the bulk of non-centrosymmetric materials, the main nonlinear process under investigation here, is given. After that, Section 7.3 formally derives the mathematical formulation of the linear GSM and extends it to nonlinear optical effects. Special attention is put on the proper discretisation of the underlying equations that allows one to use a fast iterative solver to achieve high computational efficiency. Then, in Section 7.4, the convergence and runtime

behaviour of the method are studied and compared with RCWA. The nonlinear GSM is applied to optimise the radiated second harmonic (SH) by a diffraction grating coupled to a slab waveguide in Section 7.5 before final conclusions are drawn in Section 7.6.

7.2 Rationale of the nonlinear GSM

The GSM follows a very different approach to the solution of the grating problem as compared to the RCWA. Moreover, because its precise mathematical formulation, which will be given in Section 7.3, is technically involved, this initial section describes the general solution approach of the GSM using as an illustrative example SHG in the bulk of non-centrosymmetric media, which is the main nonlinear process under investigation in this chapter.

7.2.1 Physical model for second harmonic generation

This derivation considers a quadratically nonlinear optical medium and a time-harmonic electric field, $\mathscr{E}(\mathbf{r},t)$, propagating in the medium at the fundamental frequency, ω_0,

$$\mathscr{E}(\mathbf{r},t) = \mathbf{E}(\mathbf{r})e^{-i\omega_0 t} + \text{c.c.}, \tag{7.1}$$

where $\mathbf{r}$ and t are the position vector and time, respectively. Through nonlinear interaction with the optical medium, this field gives rise to a nonlinear source polarization, $\mathscr{P}(\mathbf{r},t) = \mathbf{P}^{\mathrm{NL}}(\mathbf{r})\exp(-i\Omega t) + c.c.$, which oscillates at the SH frequency, $\Omega = 2\omega_0$. This polarization is the source of the electromagnetic field generated at the SH and is related to the excitation field via the tensor of second-order nonlinear susceptibility, $\boldsymbol{\chi}^{(2)}$:

$$P_\gamma^{\mathrm{NL}}(\mathbf{r}) = \epsilon_0 \sum_{\alpha,\beta} \chi^{(2)}_{\gamma\alpha\beta} E_\alpha(\mathbf{r}) E_\beta(\mathbf{r}). \tag{7.2}$$

In this equation and what follows, Greek indices take the values of the Cartesian coordinates x, y, and z.

Under the undepleted pump approximation, the initially nonlinear problem in the time-domain is solved in three steps: *i*) Calculate the field at the FF, ω_0; *ii*) evaluate the nonlinear polarization generated at the SH via Eq. (7.2); and *iii*) calculate the field at the SH. The first and last steps are performed using the linear and nonlinear (extended) versions of the GSM, respectively.

7.2.2 The GSM for inhomogeneous problems

At both the FF and SH, the electromagnetic field is governed by the electromagnetic wave equation with a physical source current, $\mathbf{J}^{\mathrm{ext}}(\mathbf{r})$,

$$\nabla \times \nabla \times \mathbf{E} - \omega^2 \mu_0 \epsilon \mathbf{E} = i\omega\mu_0 \mathbf{J}^{\mathrm{ext}}, \tag{7.3}$$

where for convenience the spatial dependence of the variables has been suppressed and a generic frequency, ω, has been introduced that takes the values $\omega = \omega_0$ ($\omega = \Omega$) at the FF (SH). In Eq. (7.3), $\epsilon(\mathbf{r})$ is the spatial distribution of the electric permittivity that defines the diffraction grating under consideration.

The GSM is a rigorous method for solving Eq. (7.3). It is based on the decomposition of $\epsilon(\mathbf{r})$, which defines the grating, into a simple background structure with permittivity $\epsilon_b(\mathbf{r})$ and the difference structure given by $\Delta\epsilon(\mathbf{r}) = \epsilon(\mathbf{r}) - \epsilon_b(\mathbf{r})$. The background has to be chosen such that the linear solution operator $\boldsymbol{\mathcal{N}}_b$ characterizing the background structure problem, namely the operator that associates a source $\mathbf{J}$ with the corresponding solution of Eq. (7.3) with $\epsilon = \epsilon_b$, is known. This means that

$$\mathbf{E} = \boldsymbol{\mathcal{N}}_b(\mathbf{J}) \tag{7.4}$$

solves Eq. (7.3) for any source $\mathbf{J}^{\mathrm{ext}} = \mathbf{J}$ and $\epsilon = \epsilon_b$.

In order to find the fields for an arbitrary permittivity $\epsilon(\mathbf{r})$ one rewrites Eq. (7.3) as

$$\nabla \times \nabla \times \mathbf{E} - \omega^2 \mu_0 \epsilon_b \mathbf{E} = i\omega\mu_0 \left[\mathbf{J}^{\mathrm{ext}} + \mathbf{J}^{\mathrm{gen}}(\mathbf{E})\right], \tag{7.5}$$

where the term $\mathbf{J}^{\mathrm{gen}}(\mathbf{E}) = -i\omega\Delta\epsilon\mathbf{E}$, which depends on the solution $\mathbf{E}$ itself, is called generalised source. Using the solution operator this can be rewritten as

$$\mathbf{E} = \boldsymbol{\mathcal{N}}_b\left(\mathbf{J}^{\mathrm{ext}} + \mathbf{J}^{\mathrm{gen}}(\mathbf{E})\right) = \boldsymbol{\mathcal{N}}_b\left(\mathbf{J}^{\mathrm{tot}}(\mathbf{E})\right) \tag{7.6}$$

with the total source current $\mathbf{J}^{\mathrm{tot}}(\mathbf{E}) = \mathbf{J}^{\mathrm{ext}} + \mathbf{J}^{\mathrm{gen}}(\mathbf{E})$.

The external source term $\mathbf{J}^{\mathrm{ext}}$ at the FF is chosen such that the external field, $\mathbf{E}^{\mathrm{ext}} = \boldsymbol{\mathcal{N}}_b(\mathbf{J}^{\mathrm{ext}})$, is an incident plane wave, although other choices are possible. At the SH the external source term is given in terms of the nonlinear polarization: $\mathbf{J}^{\mathrm{ext}} = -i\Omega\mathbf{P}^{\mathrm{NL}}$.

7.3 Derivation of the nonlinear generalised source method

So far, no particular assumptions about the structure under investigation have been made and thus the formulation Eq. (7.6) is very general and valid for any structure. In order to devise a numerical algorithm, several assumptions regarding the grating structure need to be made, and the special case of diffraction gratings is considered here. The derivation of the linear GSM follows that given in Ref. [2].

To this end, consider the generic 2D periodic, corrugated grating, as depicted in Fig. 7.1. It consists of the grating region $0 \leq z \leq h$, with periods Λ_1 and Λ_2 in the transverse x and y directions, respectively, and a given permittivity distribution, $\epsilon(\mathbf{r})$, which is a periodic function of the x and y coordinates and has arbitrary z-dependence. This slab is sandwiched in-between the cover and substrate, which consist of linear optical materials

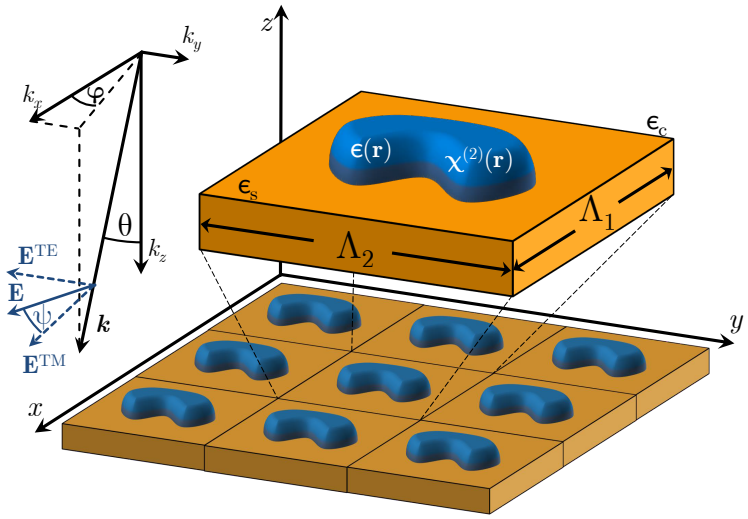

Figure 7.1: Setting for the GSM and close-up of the unit cell of a 2D corrugated grating containing nonlinear material. The grating is described by $\epsilon(\mathbf{r})$ and $\boldsymbol{\chi}^{(2)}(\mathbf{r})$ and is under plane wave incidence with wavevector $\mathbf{k}$.

with relative permittivity ϵ_c and ϵ_s, respectively. It is assumed that only the grating region contains nonlinear material, i.e. $\boldsymbol{\chi}^{(2)}(\mathbf{r}) = 0$ if $z < 0$ or $z > h$.

Due to its periodicity, the permittivity of the structure can be decomposed in a Fourier series,

$$\epsilon(\mathbf{r}) = \sum_{n=-\infty}^{\infty} \epsilon_n(z) e^{i[(n_1 K_{1x} + n_2 K_{2x})x + (n_1 K_{1y} + n_2 K_{2y})y]}, \tag{7.7}$$

where the sum over $n \equiv (n_1, n_2)$ is to be understood as a double infinite sum over the integers n_1 and n_2, $n_{1,2} \in \mathbb{Z}$, and $K_i = 2\pi/\Lambda_i$, for $i = 1, 2$. Because $\epsilon(\mathbf{r})$ is a periodic function in the transverse plane, one can use Bloch theorem to express the electric field as a Fourier series of functions that are pseudo periodic on the transverse coordinates x and y,

$$\mathbf{E}(\mathbf{r}) = \sum_{n=-\infty}^{\infty} \mathbf{E}_n(z) e^{i(k_{nx} x + k_{ny} y)}, \tag{7.8}$$

where $k_{nx/y} = k_{x/y} + n_1 K_{1x/y} + n_2 K_{2x/y}$ are the projections of the wavevector of the nth diffraction order. The principal direction of the central diffraction order, which is defined by $n_1 = n_2 = 0$, is given by $\mathbf{k} = k_c(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$, where $k_c = \omega_0\sqrt{\epsilon_0\epsilon_c\mu_0}$ is the wavenumber in the cover region. For oblique incidence, this leads to a phase-shift of the electric field $\mathbf{E}(\mathbf{r})$ of $e^{i(k_{0x}\Lambda_1 + k_{0y}\Lambda_2)}$ over a unit cell.

Since this analysis does not include nonlinear optical effects in the cover and substrate materials and since the nonlinear polarization $\mathbf{P}^{\mathrm{NL}}(\mathbf{r})$ arises due to a periodically distributed field (7.8) at the FF, both the external current $\mathbf{J}^{\mathrm{ext}} = -i\Omega\mathbf{P}^{\mathrm{NL}}$ at the SH and the generalised

source term can be expressed as Fourier series,

$$\mathbf{J}^{\mathrm{tot}} = -i\Omega\mathbf{P}^{\mathrm{NL}} - i\Omega(\epsilon - \epsilon_b)\mathbf{E} = \sum_{n=-\infty}^{\infty} \left[\mathbf{j}_n^{\mathrm{NL}}(z) + \mathbf{j}_n^{\mathrm{gen}}(z)\right] e^{i(k_{nx}x + k_{ny}y)}. \tag{7.9}$$

The coefficients $\mathbf{j}_n^{\mathrm{gen}}(z)$ depend on all field terms $\mathbf{E}_m(z)$

$$\mathbf{j}_n^{\mathrm{gen}}(z) = \quad -i\omega \left\{[\epsilon(z) - \epsilon_b(z)]\, \mathbf{E}(z)\right\}_n = -i\omega \sum_{m=-\infty}^{\infty} \left[\epsilon(z) - \epsilon_b(z)\right]_{n-m} \mathbf{E}_m(z). \tag{7.10}$$

The solution operator $\mathscr{N}_b$ for a periodic source current, $\mathbf{j}_n(z)e^{i(k_{nx}x+k_{ny}y)}$, in homogeneous space with constant permittivity ϵ_b reads [2, 4, 5]:

$$\mathbf{E}_n(z) = \frac{i}{\omega} \int \mathbf{G}_n(z, z')\mathbf{j}_n(z')dz', \tag{7.11}$$

with the tensorial Green's function

$$\begin{aligned}\mathbf{G}_n(z, z') =& \mathsf{Y}_n^{+} H(z - z')e^{ik_{nz}^b(z-z')} + \mathsf{Y}_n^{-} H(z' - z)e^{ik_{nz}^b(z'-z)} \\ &+ \epsilon_b^{-1}\mathbf{e}_z\mathbf{e}_z^T\delta(z - z')\end{aligned} \tag{7.12}$$

where the entries of the 3×3 matrix $\mathsf{Y}^{\pm}$ are given by

$$\mathsf{Y}_{m,\alpha\beta}^{\pm} = \frac{k_{n\alpha}^{\pm}k_{n\beta}^{\pm} - k_b^2\delta_{\alpha\beta}}{2\omega\epsilon_b k_{nz}^b}. \tag{7.13}$$

Here, $\mathbf{k}_n^{\pm} = (k_{nx}, k_{ny}, \pm k_{nz}^b)$ is the wavevector of the nth diffraction order in the background medium ϵ_b with the z-component

$$k_{nz}^b = \sqrt{\mu_0\epsilon_b\omega^2 - k_{nx}^2 - k_{ny}^2} \tag{7.14}$$

given by the dispersion relation for ϵ_b. One can see that the tensor $\mathbf{G}_n$ gives rise to upwards and downwards propagating plane waves in the background medium and a stationary term. To express the solution in terms of a superposition of plane waves, the contribution of the δ-term in Eq. (7.12) is eliminated by defining the modified electric field

$$\tilde{E}_\alpha(z) = E_\alpha(z) - \frac{\delta_{\alpha,z}}{i\omega\epsilon_b} j_\alpha^{\mathrm{tot}}(z). \tag{7.15}$$

The Fourier components of this modified field, $\tilde{\mathbf{E}}_n(z) = \mathsf{Q}_n\mathbf{a}_n(z)$, can be expressed via their TE/TM amplitudes, $\mathbf{a}_n(z) = \left(a_{en}^{+}(z), a_{en}^{-}(z), a_{hn}^{+}(z), a_{hn}^{-}(z)\right)$, and the matrix $\mathsf{Q}_n = (\mathbf{s}_n^{+}, \mathbf{s}_n^{-}, \mathbf{p}_n^{+}, \mathbf{p}_n^{-})$, whose entries are given by the TE/TM basis vectors from Appendix B.

Additionally, Eq. (7.15) leads to a useful property of its z-component:

$$\tilde{E}_z = \frac{D_z}{\epsilon_b}. \tag{7.16}$$

This allows the total source to be rewritten as

$$\begin{aligned}\mathbf{J}^{\mathrm{tot}} =& \mathbf{J}^{\mathrm{ext}} + \mathbf{J}^{\mathrm{gen}}(\mathbf{E}) = \mathbf{J}^{\mathrm{ext}} - i\omega(\epsilon\mathbf{E} - \epsilon_b\mathbf{E}) = \\ & - i\omega(\mathbf{P}^{\mathrm{NL}} + \epsilon\mathbf{E} - \epsilon_b\mathbf{E}) = -i\omega\left(\mathbf{D} - \mathbf{D}_b\right),\end{aligned} \tag{7.17}$$

where $\mathbf{D}_b = \epsilon_b\mathbf{E}$,

$$\mathbf{D} = \epsilon\mathbf{E} + \mathbf{P}^{\mathrm{NL}}, \tag{7.18}$$

and $\mathbf{P}^{\mathrm{NL}} = 0$ if $\omega = \omega_0$.

7.3.1 Fourier factorisation for the nonlinear GSM

So far, an *infinite* number of terms in the Fourier series of structure Eq. (7.7) and solution Eq. (7.8) was considered. However, it is well known [6–8] that the factorisation of products of periodic functions with a *finite* number of terms has to be performed with care, namely the correct approach is dictated by the continuity properties of the factors in the product, as we have seen in Section 3.3.3. To be more specific, products of functions with non-simultaneous discontinuities (e.g., the tangential component of the electric displacement field $D_\parallel = \epsilon E_\parallel + P_\parallel^{\mathrm{NL}}$) can be factorized using Laurent's product rule, whereas continuous products of functions with simultaneous discontinuities (the normal component $D_\perp = \epsilon E_\perp + P_\perp^{\mathrm{NL}}$ is continuous at interfaces, whereas ϵ and $E_\perp$ can be discontinuous) can be factorized using the inverse rule. In the context of numerical methods for periodic structures in the Fourier space, such as the RCWA and GSM, the violation of the correct factorisation rules causes slow convergence with respect to the number of retained diffraction orders. In contrast to previous work [9], where the correct factorisation was only used in the linear part of the calculations, the correct Fourier factorisation for nonlinear part of the calculations is derived here.

The derivation of this modified factorisation rule is presented before its impact on the accuracy of the calculations is investigated in Section 7.4. Thus, to illustrate why a modified factorisation rule for the inhomogeneous wave equation is necessary, the example from Section 3.3.3 and Ref. [7], is adapted to the case at hand, by defining four periodic functions on the interval $[-\pi, \pi[$:

$$f(x) = \begin{cases} a, & |x| < \pi/2, \\ a/2, & |x| \geq \pi/2, \end{cases} \tag{7.19a}$$

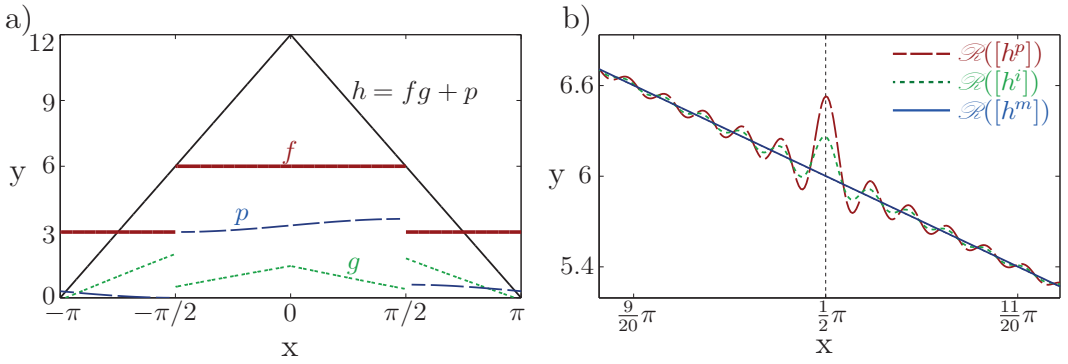

Figure 7.2: Illustration of the correct Fourier factorisation for nonlinear calculations. a) Discontinuous functions f (thick brown), g (dashed green), and p (solid blue) and continuous function h (solid black), for $a = 6$, $b = 2$, $c = 3$. b) Reconstruction of Fourier series with 200 components around discontinuity obtained by the product rule (brown), inverse rule (green), and modified inverse rule (blue).

$$p(x) = 0.1c[1 + \sin(x)] + \begin{cases} c, & |x| < \pi/2, \\ 0, & |x| \geq \pi/2, \end{cases} \tag{7.19b}$$

$$g(x) = -\frac{p(x)}{f(x)} + \begin{cases} b(1 - |x|/\pi), & |x| < \pi/2, \\ 2b(1 - |x|/\pi), & |x| \geq \pi/2, \end{cases} \tag{7.19c}$$

$$h(x) = f(x)g(x) + p(x). \tag{7.19d}$$

Here, $a, b > 0$ and $c \geq 0$ are arbitrary constants. One can easily see that the function d is continuous, whereas f and g have concurrent discontinuities at $|x| = \pi/2$. For $c \neq 0$, p is also discontinuous at $|x| = \pi/2$. Otherwise, Li's example [7] is obtained. Fig. 7.2(a) shows the functions for $a = 6$, $b = 2$, and $c = 3$.

In order to obtain the vector, $[h]$, of Fourier coefficients of h, three different factorisations are compared, the results being depicted in Fig. 7.2(b). The product rule yields $[h^p] = [\![f]\!][g] + [p]$, where $[\![f]\!]$ is the Toeplitz matrix of the Fourier coefficients of f. The reconstruction, $\mathscr{R}([d^p])$, of the function h shows strong oscillations at $|x| = \pi/2$ around the expectedly continuous function h itself. The reason for this behaviour is that the product rule propagates the overshoot from $\mathscr{R}([f])$ and $\mathscr{R}([g])$ to the reconstruction of the product $\mathscr{R}([\![f]\!][g])$. In addition, the oscillatory behaviour of $\mathscr{R}([p])$ at $|x| = \pi/2$ adds to the variations of the reconstructed product, $\mathscr{R}([h^p])$.

A direct application of the inverse rule for the product yields $[h^i] = [\![1/f]\!]^{-1}[g] + [p]$. The oscillations of the reconstruction $\mathscr{R}([h^i])$ are now less pronounced but still consist of additive contributions from both terms $\mathscr{R}\left([\![1/f]\!]^{-1}[g]\right)$ and $\mathscr{R}([p])$.

A modification of the inverse rule is proposed here: to this end, the continuous function $h(x) = f(x)g(x) + p(x) = f(x)\big[g(x) + p(x)/f(x)\big]$ is rewritten as a product of discontinuous functions *in the space-domain*. The two factors have complementary dis-

continuities, whence the inverse rule applied to this product yields the correct factorisation, $[h^m] = [\![1/f]\!]^{-1}([g] + [p/f])$. This is illustrated by the reconstruction $\mathscr{R}([h^m])$, which shows no spurious oscillations at the points of discontinuity.

The next sections will incorporate this modified factorisation rule into the nonlinear GSM and describe the discretisation of the computational quantities.

7.3.2 Discretisation of the EM fields by Fourier series

Armed with this correct factorisation rule, let us now consider the Fourier factorisation of Eq. (7.17), but note first that all terms in this equation implicitly have a dependence on the three spatial variables, i.e. $\mathbf{D}(x,y,z) = \epsilon(x,y,z)\mathbf{E}(x,y,z) + \mathbf{P}^{\mathrm{NL}}(x,y,z)$. This section will explain the discretisation of Eq. (7.17) in the transverse directions x and y by the Fourier series description Eq. (7.8) before the next section will describe the discretisation in the propagation direction z.

In the context of the nonlinear GSM, h, f, g, and p from last section's example correspond to the continuous normal component of the displacement field, $D_\perp$, the permittivity, ϵ, $E_\perp$, and $P_\perp^{\mathrm{NL}}$, respectively. The correct factorisation of tangential and parallel components of $\mathbf{D}$ therefore reads at a fixed altitude z:

$$[D_\parallel(z)] = [\![\epsilon(z)]\!][E_\parallel(z)] + [P_\parallel^{\mathrm{NL}}(z)], \tag{7.20a}$$

$$[D_\perp(z)] = [\![1/\epsilon(z)]\!]^{-1}\left([E_\perp(z)] + [P_\perp^{\mathrm{NL}}(z)/\epsilon(z)]\right). \tag{7.20b}$$

It should be specified again that $[f]$ denotes the vector of $N_0 = (2N_1+1)(2N_2+1)$ Fourier coefficients of a scalar function f. The truncation of Fourier terms with $n_1 = -N_1, \ldots, N_1$ and $n_2 = -N_2, \ldots, N_2$, with N_1 and N_2 integers, corresponds to a rectangular truncation pattern in the Fourier space. Accordingly, the Fourier coefficients of the total source current, $\mathbf{j}_n = \mathbf{j}_n^{\mathrm{NL}} + \mathbf{j}_n^{\mathrm{gen}}$, defined by Eq. (7.9) are given by

$$[j_\parallel] = -i\omega\epsilon_b\left\{\left(\left[\!\!\left[\frac{\epsilon}{\epsilon_b}\right]\!\!\right] - [\![1]\!]\right)[E_\parallel] + \epsilon_b^{-1}[P_\parallel^{\mathrm{NL}}]\right\}, \tag{7.21a}$$

$$[j_\perp] = -i\omega\epsilon_b\left\{\left(\left[\!\!\left[\frac{\epsilon_b}{\epsilon}\right]\!\!\right]^{-1} - [\![1]\!]\right)[E_\perp] + \left[\!\!\left[\frac{\epsilon_b}{\epsilon}\right]\!\!\right]^{-1}[P_\perp^{\mathrm{NL}}/\epsilon]\right\}. \tag{7.21b}$$

To simplify the notation, the dependence of all Fourier series coefficient on the coordinate z is to be dropped unless otherwise specified.

It remains to reformulate these equations, which are given in terms of the $(\perp,\parallel)$-components of $\mathbf{E}$, in terms of the Cartesian components of $\tilde{\mathbf{E}}$, the latter ones being the actual variables used in the GSM formalism. To this end, a local normal/tangential vector

field is defined so that for any vectorial quantity $\mathbf{v}$ we can write:

$$\begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} = \mathsf{B} \begin{pmatrix} v_n \\ v_\psi \\ v_\phi \end{pmatrix}, \tag{7.22}$$

where the orthogonal transformation matrix

$$\mathsf{B} = \left(\hat{\mathbf{n}}, \hat{\psi}, \hat{\phi}\right) = \begin{pmatrix} \cos\phi\sin\psi & \cos\phi\cos\psi & -\sin\phi \\ \sin\phi\sin\psi & \sin\phi\cos\psi & \cos\phi \\ \cos\psi & -\sin\psi & 0 \end{pmatrix} \tag{7.23}$$

is a concatenation of the unit vectors in normal, and both tangential directions $\hat{\mathbf{n}}$, $\hat{\psi}$, and $\hat{\phi}$, respectively.

Using this transformation one obtains analogously to the procedure described in Appendix A in [2]:

$$[j_\alpha] = -i\omega\epsilon_b \sum_\beta \left(\Delta\delta_{\alpha\beta} - \mathsf{D}\Gamma_{\alpha\beta}\right)[E_\beta] - i\omega\epsilon_b[\bar{p}^p_\alpha], \tag{7.24}$$

where the vectors of Fourier series coefficients of the intermediate polarization $\bar{p}^p$ are given by

$$\begin{pmatrix} [\bar{p}^p_x] \\ [\bar{p}^p_y] \\ [\bar{p}^p_z] \end{pmatrix} = \mathsf{B}\,\mathrm{diag}\left(\left[\!\left[\frac{\epsilon_b}{\epsilon}\right]\!\right]^{-1}, \epsilon_b^{-1}, \epsilon_b^{-1}\right) \begin{pmatrix} [P^{\mathrm{NL}}_n/\epsilon] \\ [P^{\mathrm{NL}}_\psi] \\ [P^{\mathrm{NL}}_\phi] \end{pmatrix}. \tag{7.25}$$

The matrices Δ and D are defined as

$$\Delta = \left[\!\left[\frac{\epsilon}{\epsilon_b}\right]\!\right] - I \quad \text{and} \quad \mathsf{D} = \left[\!\left[\frac{\epsilon}{\epsilon_b}\right]\!\right] - \left[\!\left[\frac{\epsilon_b}{\epsilon}\right]\!\right]^{-1} \tag{7.26}$$

and $\Gamma_{\alpha\beta}$ are the Fourier images of the entries of the matrix

$$\Gamma(z) = \begin{pmatrix} \cos^2\phi\sin^2\psi & \sin\phi\cos\phi\sin^2\psi & \cos\phi\sin\psi\cos\psi \\ \sin\phi\cos\phi\sin^2\psi & \sin^2\phi\sin^2\psi & \sin\phi\sin\psi\cos\psi \\ \cos\phi\sin\psi\cos\psi & \sin\phi\sin\psi\cos\psi & \cos^2\psi \end{pmatrix}. \tag{7.27}$$

Note that with this definition of Γ, Eq. (7.24) implements the one-sided NVF approach to Fourier factorisation from Eq. (3.41a).

In order to eliminate the physical electric field $\mathbf{E}$ in favour of the computational field

$\tilde{\mathbf{E}}$ one inserts

$$\frac{[j_z]}{i\omega\epsilon_b} = \mathsf{D}\left(\Gamma_{zx}[E_x] + \Gamma_{zy}[E_y]\right) + \left(\mathsf{D}\Gamma_{zz} - \Delta\right)[E_z] - [\bar{p}_z^p] \tag{7.28}$$

into the definition (7.15) of $\tilde{\mathbf{E}}$ to obtain

$$[\tilde{E}_z] = -\mathsf{D}\left(\Gamma_{zx}[E_x] + \Gamma_{zy}[E_y]\right) + \mathsf{C}[E_z] + [\bar{p}_z^p], \tag{7.29}$$

where $\mathsf{C} = \Delta - \mathsf{D}\Gamma_{zz}$. This yields $[\mathbf{E}]$ resolved in terms of $[\tilde{\mathbf{E}}]$. This is expressed in matrix form as

$$\begin{pmatrix} [E_x] \\ [E_y] \\ [E_z] \end{pmatrix} = \begin{pmatrix} \mathsf{I} & 0 & 0 \\ 0 & \mathsf{I} & 0 \\ \mathsf{C}^{-1}\mathsf{D}\Gamma_{zx} & \mathsf{C}^{-1}\mathsf{D}\Gamma_{zy} & \mathsf{C}^{-1} \end{pmatrix} \begin{pmatrix} [\tilde{E}_x] \\ [\tilde{E}_y] \\ [\tilde{E}_z] \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ \mathsf{C}^{-1}[\bar{p}_z^p] \end{pmatrix}. \tag{7.30}$$

Plugging this formula back into Eq. (7.24) provides the total source current, which consists of the generalised source dependent on the unknown field $\tilde{\mathbf{E}}$ and a physical source whose origin is the known polarization:

$$\frac{[j_\alpha]}{i\omega\epsilon_b} = \sum_\beta \mathsf{W}_{\alpha\beta}[\tilde{E}_\beta] + [\tilde{p}_\alpha^p], \tag{7.31}$$

where the modified polarization reads

$$[\tilde{p}_\alpha^p] = -[\bar{p}_\alpha^p] + \left(\Delta\delta_{\alpha z} - \mathsf{D}\Gamma_{\alpha z}\right)\mathsf{C}^{-1}[\bar{p}_z^p] \tag{7.32}$$

and the matrix W is defined as

$$\mathsf{W} = \begin{pmatrix} \mathsf{N}_{xx} + \mathsf{D}\Gamma_{xx} - \Delta & \mathsf{N}_{xy} + \mathsf{D}\Gamma_{xy} & \mathsf{D}\Gamma_{xz}\mathsf{C}^{-1} \\ \mathsf{N}_{yx} + \mathsf{D}\Gamma_{xx} & \mathsf{N}_{xy} + \mathsf{D}\Gamma_{yy} - \Delta & \mathsf{D}\Gamma_{yz}\mathsf{C}^{-1} \\ -\mathsf{C}^{-1}\mathsf{D}\Gamma_{zx} & -\mathsf{C}^{-1}\mathsf{D}\Gamma_{zy} & \mathsf{C}^{-1} - \mathsf{I} \end{pmatrix}. \tag{7.33}$$

Here, $\mathsf{N}_{\alpha\beta} = \mathsf{D}\Gamma_{\alpha z}\mathsf{C}^{-1}\mathsf{D}\Gamma_{z\beta}$ combines the geometrical normal field information contained in Γ with the physical structure given by ϵ.

This concludes the discretisation of the physical quantities in the transverse direction. Note that all matrices and vectors of Fourier series coefficients throughout this section, $[E_\alpha(z)]$, $[\tilde{p}_\alpha^p(z)]$, $\left[\!\left[\frac{\epsilon(z)}{\epsilon_b(z)}\right]\!\right]$, $W(z)$ etc. still possess an arbitrary dependence on the z-direction which will be resolved in the next section.

7.3.3 Derivation and solution of the integral equation in the nonlinear GSM

The combination of the main results obtained so far, namely Eq. (7.31) that expresses the source in terms of $\tilde{\mathbf{E}}$ and the Green's tensor Eq. (7.11), yields a system of Fredholm integral equations of the second kind for the unknown amplitudes,

$$\mathbf{a}_n(z) = \mathbf{a}_n^0(z) + \int_{z'=0}^{h} \mathsf{R}_n(z,z')\bar{\mathsf{Q}}_n \sum_{m=-N_0}^{N_0} \sum_{\beta} \mathsf{W}_{\alpha\beta,nm}(z')\mathsf{Q}_m \mathbf{a}_m(z')dz', \tag{7.34}$$

for $n = 1,\ldots,N_0$, where the sum over $m \equiv (m_1, m_2)$ is to be understood as a double finite sum over the integers $m_1 = -N_1,\ldots,N_1$ and $m_2 = -N_2,\ldots,N_2$. The matrix $\mathsf{R}_n(z,z') \in \mathbb{C}^{4\times4}$ incorporates the propagation of TE- and TM-polarized, upward and downward plane wave amplitudes of the n^{th} Fourier component inside the slab-background structure and their reflections at the top and bottom interfaces (see Eqs. (61) and (62) in [2] for the exact formula). The matrix $\bar{\mathsf{Q}}_n$ transforms the field amplitudes, $\tilde{\mathbf{E}}_n(z)$, back to the coefficients, $\mathbf{a}_n(z) = \bar{\mathsf{Q}}_n\tilde{\mathbf{E}}_n(z)$, and is given by

$$\bar{\mathsf{Q}}_n = \frac{1}{2}\begin{pmatrix} \frac{k_{ny}}{\gamma_n} & -\frac{k_{nx}}{\gamma_n} & 0 \\ \frac{k_{ny}}{\gamma_n} & -\frac{k_{nx}}{\gamma_n} & 0 \\ -\frac{\omega\epsilon_b k_{nx}}{k_{nz}\gamma_n} & -\frac{\omega\epsilon_b k_{ny}}{k_{nz}\gamma_n} & \frac{\omega\epsilon_b}{\gamma_n} \\ \frac{\omega\epsilon_b k_{nx}}{k_{nz}\gamma_n} & \frac{\omega\epsilon_b k_{ny}}{k_{nz}\gamma_n} & \frac{\omega\epsilon_b}{\gamma_n} \end{pmatrix} \tag{7.35}$$

The known values of $\mathbf{a}_n^0(z)$ are either the amplitudes of the incident plane wave electric field in the background structure at the FF or the amplitudes of the external source field $\mathscr{N}_b(\mathbf{J}^{\text{ext}})$ at the SH; the latter are given by

$$\mathbf{a}_n^0(z) = \int_{z'=0}^{h} \mathsf{R}_n(z,z')\bar{\mathsf{Q}}_n\tilde{\mathbf{p}}_n^p(z')dz'. \tag{7.36}$$

One can easily see that the modified factorisation rule does not affect the generalised source but only changes the external polarization. This means, that all algorithmic enhancements, that make the linear GSM an efficient numerical method, can be applied for the nonlinear version of the GSM.

Using the midpoint rule, the system Eq. (7.34) of integral equations is discretised at N_L equidistant points, $z_p = (p-0.5)h/N_L$, $p = 1,\ldots,N_l$. This results in a system of linear equations,

$$\mathbf{a}_{np} = \mathbf{a}_{np}^0 + \frac{h}{N_L}\sum_{q=1}^{N_L} \mathsf{R}_n(z_p,z_q)\bar{\mathsf{Q}}_n \sum_{m=-N_0}^{N_0} \sum_{\beta} \mathsf{W}_{\alpha\beta,nm}(z_q)\mathsf{Q}_m \mathbf{a}_{mq}, \tag{7.37}$$

for $n = 1,\ldots,N_0, p = 1,\ldots,N_L$. The total number of unknowns in Eq. (7.37) is $4N_0N_L$

and usually becomes prohibitively large for direct solvers to be used in practical situations of interest.

In order to take full advantage of an iterative solver, the system (7.37) is reformulated to (see also [2]):

$$\mathbf{a} = \left(\mathsf{I} + \mathsf{R}\bar{\mathsf{Q}}\mathsf{U}\mathsf{A}^{-1}\mathsf{Q}\right)\mathbf{a}^0, \tag{7.38}$$

where $\mathbf{a}$ and $\mathbf{a}^0$ are the vectors containing the discrete unknowns, $\mathbf{a}_{np} = \mathbf{a}_n(z_p)$, and the known amplitudes, $\mathbf{a}^0_{np} = \mathbf{a}^0_n(z_p)$, respectively. $\mathsf{A} = \mathsf{M} - \mathsf{Q}\mathsf{R}\bar{\mathsf{Q}}\mathsf{U}$, and $\bar{\mathsf{Q}}$ and Q are block-diagonal matrices with $\bar{\mathsf{Q}}_n$ and Q_n on their block-diagonal, respectively. The matrix $\mathsf{R} \in \mathbb{C}^{4N_0N_L \times 4N_0N_L}$ corresponds to $\mathsf{R}_n(z_p, z_q)$: Its $\big(4(n-1)N_L + 4(p-1) + [1,..,4],\ 4(n-1)N_L + 4(q-1) + [1,..,4]\big)$-entries are given by $\mathsf{R}_n(z_p, z_q)$. Hence, R is diagonal with respect to the Fourier component n and has Toeplitz-structure with respect to the layer indices p and q. The matrices

$$\mathsf{M} = \begin{pmatrix} \left[\!\left[\frac{\epsilon_b}{\epsilon}\right]\!\right]\left[\!\left[\frac{\epsilon}{\epsilon_b}\right]\!\right] & 0 & 0 \\ 0 & \left[\!\left[\frac{\epsilon_b}{\epsilon}\right]\!\right]\left[\!\left[\frac{\epsilon}{\epsilon_b}\right]\!\right] & 0 \\ 0 & 0 & \left[\!\left[\frac{\epsilon}{\epsilon_b}\right]\!\right]\left[\!\left[\frac{\epsilon_b}{\epsilon}\right]\!\right] \end{pmatrix} \sin^2\psi + \cos^2\psi, \tag{7.39}$$

$$\mathsf{F} = \mathsf{I} - \left[\!\left[\frac{\epsilon_b}{\epsilon}\right]\!\right]\left[\!\left[\frac{\epsilon}{\epsilon_b}\right]\!\right], \qquad \mathsf{G} = \mathsf{I} - \left[\!\left[\frac{\epsilon}{\epsilon_b}\right]\!\right]\left[\!\left[\frac{\epsilon_b}{\epsilon}\right]\!\right], \tag{7.40}$$

$$\mathsf{U} = \begin{pmatrix} \Delta\mathsf{M}_{xx} + \mathsf{G}\left[\!\left[\frac{\epsilon}{\epsilon_b}\right]\!\right]\Gamma_{xx} & \mathsf{G}\left[\!\left[\frac{\epsilon}{\epsilon_b}\right]\!\right]\Gamma_{xy} & \mathsf{G}\Gamma_{xz} \\ \mathsf{G}\left[\!\left[\frac{\epsilon}{\epsilon_b}\right]\!\right]\Gamma_{yx} & \Delta\mathsf{M}_{yy} + \mathsf{G}\left[\!\left[\frac{\epsilon}{\epsilon_b}\right]\!\right]\Gamma_{yy} & \mathsf{G}\Gamma_{yz} \\ \mathsf{F}\Gamma_{zx} & \mathsf{F}\Gamma_{zy} & \mathsf{M}_{zz} - \left[\!\left[\frac{\epsilon_b}{\epsilon}\right]\!\right] \end{pmatrix} \tag{7.41}$$

are defined such that $\mathsf{W} = \mathsf{U}\mathsf{M}^{-1}$ but they do not contain inverted submatrices themselves. Additionally, U and M are block-Toeplitz-Toeplitz matrices with respect to the index pair $n = (n_1, n_2)$. Due to the Toeplitz-property of its submatrices, multiplication of a vector $\mathbf{b}$ of size $3N_0N_L$ with the $3N_0N_L \times 3N_0N_L$ system matrix A can be performed in $\mathcal{O}(N_0N_L \log(N_0)\log(N_L))$ operations instead of $\mathcal{O}(N_0^2N_L^2)$ for the standard matrix-vector multiplication, see Appendix C or Ref. [10].

This allows the efficient use of iterative solvers for systems of linear equations which are based on Matrix-vector multiplication. More specifically, the generalised minimum residual method (GMRES, [11]) is used here to solve Eq. (7.38).

7.4 Numerical analysis and convergence studies

In this section the convergence and runtime characteristics of the linear and nonlinear parts of the GSM are investigated using cylindrical gratings as test examples made either of dielectric or metallic materials. Importantly, the impact of application of the correct factori-

sation rule is illustrated and compared to the conventional factorisation.

The unit cell of the grating under consideration is assumed to have a quadratic cross section with the two periods set to equal $\Lambda_1 = \Lambda_2 = \Lambda = 2\,\mu\mathrm{m}$. It contains a cylindrical disk of height, $h = 200\,\mathrm{nm}$, and radius, $r = 0.25\Lambda = 500\,\mathrm{nm}$ as per Fig. 7.3(a). The disk and substrate have refractive index, $n_s = 2$, whereas the cover region is vacuum with $n_c = 1$. A diagonal nonlinear susceptibility tensor, $\chi^{(2)}_{jkl} = \bar{\chi}\delta_{jk}\delta_{kl}$, with $\bar{\chi} = 10^{-8}\,\mathrm{m\,V^{-1}}$ is assumed for the cylinder. The incident plane wave at the fundamental wavelength of $\lambda_{\mathrm{FF}} = 1.5\,\mu\mathrm{m}$, is specified by the angles $\theta = 30°$, $\varphi = 70°$, and a mixed polarization, $\psi = 45°$.

The optical response of the device at the FF and SH is computed by the linear and nonlinear GSM, respectively. Simulations are performed for an increasing number of diffraction orders $N_0 = (2N_1 + 1)(2N_2 + 1)$ for $N_1 = N_2 \in \{3, 5, ..., 23\}$ and increasing number of layers, $N_L \in \{12, 25, 50, ..., 400\}$. In order to assess the efficiency of the algorithm, the following quantities are calculated: the reflection and transmission coefficients, $R^{FF/SH}$ and $T^{FF/SH}$, respectively, at the FF/SH. Additionally, the number $N^{\mathrm{FF}}_{\mathrm{iter}}$ and $N^{\mathrm{SH}}_{\mathrm{iter}}$ of GMRES iterations necessary to solve Eq. (7.38) at the FF and SH, respectively, and the GSM runtime needed to complete the linear (t^{FF}) and nonlinear (t^{SH}) part of the computations are recorded.

For the validation of the linear calculations, the results are compared to results from the RCWA in the formulation of Chapter 3 with NVF factorisation approach. To this end, the nonlinear GSM was implemented in MATLAB [12]. This computer software uses the same computational input and output system as the secondary RCWA implementation from Chapter 6, further facilitating this comparison of linear numerical results. Moreover, since both numerical methods have been implemented in MATLAB, a proper comparison of simulation times is possible and all simulations have been performed on a single core of a $2.5\,\mathrm{GHz}$ processor.

The results of this analysis are summarized in Figs. 7.3(b)-7.3(d). The dependence of the transmission coefficient at the FF, T^{FF}, on the number of diffraction orders, computed by using the GSM for several values of the numbers of layers N_L in which the grating is divided, is presented in Fig. 7.3(b). For comparison the same dependence determined by using the RCWA is presented. Both methods show a similar convergence pattern with respect to N_0 and the computed transmission coefficient rapidly approaches the asymptotic value of $T^{\mathrm{FF}} = 0.917772$. For the investigated lossless device the GSM conserves the incoming optical power up to a relative error of 10^{-5} for low $N_0 < 1000$ and the power conservation improves to 10^{-6} as N_0 further increases.

The inset in Fig. 7.3(b) illustrates the error in the transmission due to the discretisation of the integral equation, defined as $\Delta T = T^{\mathrm{FF}} - T^{\mathrm{FF}}_{\mathrm{RCWA}}$, calculated for several values of

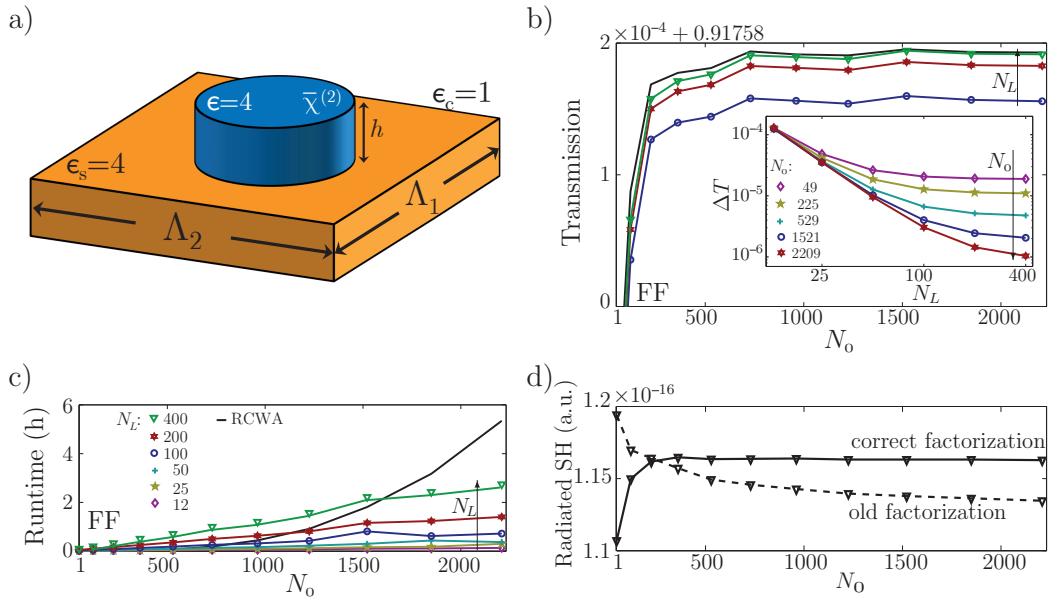

Figure 7.3: a) Cylindrical grating for the convergence test. b) Transmission efficiency obtained using GSM for increasing N_L [colours, legend from b)] and RCWA (black). Inset: discretisation error of GSM vs. N_0. c) Runtime of GSM for increasing N_L (colours) and of RCWA (black). d) Convergence behaviour of radiated SH in the direction of the incoming plane wave using the two different factorisation rules.

orders, $N_0 \in \{49, ..., 2209\}$, and for increasing N_L. Given sufficiently many orders, e.g. $N_0 = 2209$, quadratic decrease of ΔT is achieved, which is to be expected since using the midpoint rule in the discretisation of the integral equation implies quadratic convergence. For smaller N_0, e.g. $N_0 = 225$, the convergence plateaus at a small number of layers, in this case $N_L = 100$. This is because the reformulated linear system Eq. (7.38), which enables the use of the accelerated matrix-vector multiplication, is only N_0-asymptotically equivalent to the original system (7.37).

Figure 7.3(c) depicts the runtime t^{FF} necessary to compute the solution of the problem at the FF for increasing N_0 and for different N_L. The linear GSM exhibits nearly optimal runtime as t^{FF} increases log-linearly with both N_0 and N_L. This is because t^{FF} is chiefly determined by the computational work needed to iteratively find the solution of Eq. (7.38) using GMRES, which in turn depends on the number of iterations $N_{\mathrm{iter}}^{\mathrm{FF}}$ and, if $N_{\mathrm{iter}}^{\mathrm{FF}} \ll N_0 N_L$, on the time required for the system-matrix vector multiplication; this latter quantity is of order $\mathcal{O}(N_0 N_L \log(N_0) \log(N_L))$. For the investigated device, only $N_{\mathrm{iter}}^{\mathrm{FF}} = 39$ GMRES iterations are necessary to solve Eq. (7.38), where $N_{\mathrm{iter}}^{\mathrm{FF}}$ is nearly independent of N_0 and N_L (see also Table 7.1, second row). The black line in Fig. 7.3(c) indicates the runtime of the RCWA, $t_{\mathrm{RCWA}}^{\mathrm{FF}}$, which follows a $\mathcal{O}(N_0^3)$ dependence. This is the expected behaviour as the RCWA involves finding the solution of an eigenproblem of size N_0 which was was already discussed in Section 5.3.5. The asymptotic advantage of the

GSM becomes evident, e.g. at $N_L = 200$, where for $N_0 = 1225$, $t^{\mathrm{FF}}_{\mathrm{RCWA}}$ becomes larger than t^{FF}.

For the computations at the SH, the GSM computational time still follows the $\mathcal{O}(N_0 N_L \log(N_0) \log(N_L))$ trend; however, as compared to the FF, approximately three times more iterations, $N^{\mathrm{SH}}_{\mathrm{iter}} = 100$, are necessary to solve Eq. (7.38) at the SH. Two possible reasons for this behaviour can be identified: First, it was found that for plane wave excitation, the smaller the wavelength the more iterations were necessary for GMRES to solve Eq. (7.38). Second, both the r.h.s. of this equation and its solution are more inhomogeneous at the SH as compared to the FF.

The results for the generated SH, computed with $N_L = 400$, are depicted in Fig. 7.3(d). Specifically, this figure shows the radiated power at the SH in the direction of the incoming wave, normalized to the incident power at the FF. The data corresponding to the solid line was obtained by using the modified inverse rule for the source polarization and shows remarkably fast convergence to a value of $T^{\mathrm{SH}} = 1.163 \cdot 10^{-16}$ with increasing N_0. By contrast, when using the unmodified rule T^{SH} does not reach its asymptotic value even for as many as $N_0 = 2209$ orders, despite the fact that the FF solution is already converged if $N_0 \gtrsim 750$ as per see Fig. 7.3(b). Note that the old and correct factorisations seem to tend slightly different values. This can however hardly be decided, as the old factorisation is far from reaching convergence. Similar behaviour was found in the example from Section 3.6 or Ref. [8], the numerical results obtained by a method without correct Fourier factorisation converge very slowly, and might appear to reach a different value. Therefore, Fig. 7.3(d) indicates that the slow convergence due to incorrect Fourier factorisation carries over into nonlinear calculations.

The convergence and runtime properties of the GSM where investigated when it was applied to simulate diffraction gratings made of three different lossless dielectric materials, a lossy dielectric, and metallic components. The results, shown in Table 7.1, provide valuable insights into the characteristics of the GSM. Thus, clear trends can be observed in the case of lossless dielectric gratings: the higher the refractive index n is the more GMRES

n	$N^{\mathrm{FF}}_{\mathrm{iter}}$	$N^{\mathrm{SH}}_{\mathrm{iter}}$	T^{FF}	A^{FF}	T^{SH}
1.45	21	38	0.972998	$2.6 \cdot 10^{-8}$	$2.88 \cdot 10^{-17}$
2	39	100	0.917772	$8.0 \cdot 10^{-7}$	$1.16 \cdot 10^{-17}$
3	105	400	0.825020	$5.8 \cdot 10^{-6}$	$4.57 \cdot 10^{-18}$
$1.5{+}1i$	96	157	0.620149	0.283418	$7.41 \cdot 10^{-19}$
$0.36{+}6.5i$	2160	1968	0.381113	0.193255	$5.99 \cdot 10^{-18}$

Table 7.1: Characteristics of the nonlinear GSM for dielectric and lossy materials

iterations $N_{\text{iter}}^{\text{FF}}$ are necessary at the FF and the number of iterations $N_{\text{iter}}^{\text{SH}}$ at the SH increases faster than $N_{\text{iter}}^{\text{FF}}$. The numerical absorption, $A^{\text{FF}} = 1 - T^{\text{FF}} - R^{\text{FF}}$, which in the case of gratings made of lossless materials quantifies the degree to which the algorithm conserves the energy, decreases with the index of refraction. The lossy material with $n = 1.5 + i$ can be efficiently analysed by the GSM: the $N_{\text{iter}}^{\text{FF}}$ is relatively small and increases only by 50% for the calculation of the SH electromagnetic field.

The metallic grating requires a more detailed discussion. The derivation of the GSM formally allows complex values of the permittivity $\epsilon(\mathbf{r})$ and the implementation of the algorithm does indeed show that the method can be applied to simulate diffraction by lossy structures (see the $n = 1.5 + i$ example in Table 7.1). However, the very large number of required GMRES iterations makes the method particularly slow when applied to structures containing very lossy components, namely a material with $n = 0.36 + 6.5i$. An additional finding is that N_{iter} for metals increases strongly when N_0 increases: the presented results for $n = 0.36 + 6.5i$ are for $N_1 = N_2 = 11$; increasing N_0 would require $N_{\text{iter}} > 5000$ to achieve convergence of GMRES. Summing up these results, the GSM in the presented form is most efficient and most suitable for shallow, low refractive index contrast, dielectric devices. However, an improvement of the GSM using curvilinear coordinates [3] allows one to efficiently calculate metallic structures as well. This this version of the algorithm was however not adapted to the nonlinear case here.

7.5 Application to SHG in textured slab waveguides

In this section the versatility of the nonlinear GSM is illustrated by applying it to a problem of practical importance, namely to the maximization of SHG. Because of the weak nature of nonlinear optical interactions, the radiated SH in practical settings is orders of magnitude smaller than the linear excitation. Therefore, field enhancing mechanisms are particularly important in this context as they can be employed to increase the efficiency of SHG in applications. To this end, SHG in a diffraction grating resonantly coupled to a slab waveguide is considered here, both optical devices being made of quadratically nonlinear optical materials.

To add specificity to this setting problem, it is assumed that the nonlinear material in the device is GaAs (refractive index $n = 3.4$), which due to its large optical nonlinearity is a particularly suitable choice for high SHG in device applications. It crystallizes in the zincblende lattice type, so that it belongs to the crystal point group $\bar{4}3m$. Hence, the non-vanishing components of its second-order susceptibility tensor are $\chi^{(2)}_{xyz} = \chi^{(2)}_{xyz} = \chi^{(2)}_{yxz} = \chi^{(2)}_{yzx} = \chi^{(2)}_{zxy} = \chi^{(2)}_{zyx} = 7.4 \times 10^{-10}\,\text{m}\,\text{V}^{-1}$ [13, 14], which is a relatively large value as compared to that of most other non-centrosymmetric materials. CaF_2 ($n_s = 1.4$) is chosen as substrate and the cover is considered to be air ($n_c = 1$), which provides the high refractive index contrast needed to achieve good field confinement in the slab waveguide.

In order for the incident plane waves to couple and interact with the slab waveguide modes a periodic optical grating that breaks the translational symmetry of the structure is placed on top of the waveguide. As a result the propagation constant, $\beta_\nu(\lambda)$, of the νth slab waveguide mode is folded within the first Brillouin zone of the grating. When the difference between the tangential component of the incident wavevector, $\mathbf{k}_\parallel$, and the propagation constant $\beta_\nu(\lambda)$ is equal to a multiple of the grating vector this waveguide mode can be resonantly excited, as we have already seen theoretically through Eq. (2.43) and practically in Section 6.6. In particular, one expects to observe strong field enhancement in the slab waveguide at the corresponding resonance wavelengths and consequently significantly stronger nonlinear optical interaction, an effect that has been recently used to demonstrate enhanced optical absorption in such optical devices [15].

In the context of SHG there are two kinds of resonances that can affect the nonlinear optical response of the device: *linear resonances* correspond to the grating-induced, resonant excitation of a waveguide mode at the FF by a plane wave incident at a certain angle and wavelength, λ_{FF}. One can expect that the optical field at the FF is strongly enhanced and confined inside the waveguide, thus a stronger nonlinear polarization is created, which in turn excites a stronger field at the SH wavelength, $\lambda_{\mathrm{SH}} = \lambda_{\mathrm{FF}}/2$. This phenomenon of direct enhancement of the nonlinear response at the SH due to the excitation of a linear resonance at the FF will be called an *inherited resonance* at λ_{SH}. The second type of nonlinear resonances are *intrinsic resonances*, which are observed when a waveguide mode is excited at the SH wavelength by the dipoles associated to the nonlinear polarization. In this case the field at the SH is confined inside the waveguide and therefore shows remarkably large radiated SH intensity. Note that the intrinsic resonances are not necessarily grating coupled with the radiative continuum and therefore they can be viewed as nonlinear dark modes of the system [16]. In particular, one expects that these resonances have large quality factor. In summary, the inherited and intrinsic resonances are called *nonlinear* resonances, since they are observed at the SH. For the sake of simplicity, the modes causing the linear and nonlinear resonances are denoted as linear and nonlinear modes. The SH wavelengths λ_{SH} for which both types of nonlinear modes are excited simultaneously are of particular interest because at these wavelengths, this doubly resonant mechanism of SHG, can lead to a remarkably large nonlinear optical response of the device. In the remaining of this section, these ideas are illustrated by studying the nonlinear optical response of 1D- and 2D-periodic implementations of these devices.

7.5.1 One-dimensional binary gratings

Initially, a 1D grating is considered. Although having a simple structure, it already provides us the necessary insights to completely understand the more complex 2D design. Thus, consider the 1D binary GaAs grating on top of a GaAs slab waveguide in Fig. 7.4(a), which

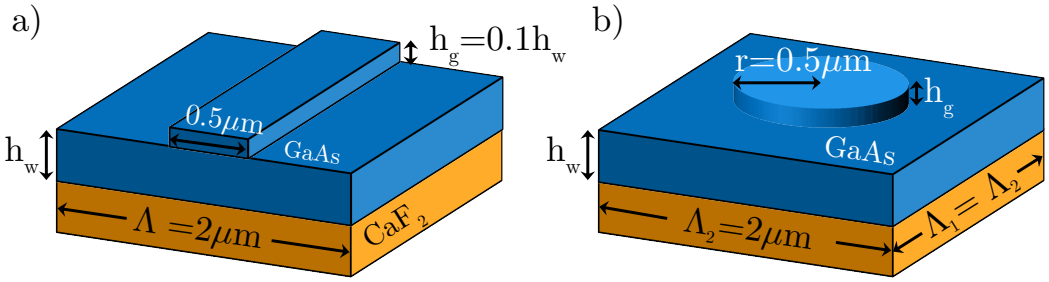

Figure 7.4: a) Unit cell of a one-dimensional binary grating and b) two-dimensional cylindrical grating on top of a slab waveguide made of nonlinear GaAs on a CaF_2 substrate.

is placed onto a CaF_2 substrate. The height of the slab waveguide, h_w, is a free design parameter and varies from $0.25\,\mu\mathrm{m}$ to $0.5\,\mu\mathrm{m}$. The period of the grating with filling factor $\rho = 0.25$ is $\Lambda = 2\,\mu\mathrm{m}$, the height of the grating region is chosen relative to the height of the slab waveguide, $h_g = 0.1h_w$.

At the FF, TM polarized plane waves with wavelength λ_{FF} ranging from $3.6\,\mu\mathrm{m}$ to $4.4\,\mu\mathrm{m}$ are incident onto the device under a $\theta = 30°$ angle. By considering the slab waveguide region as homogeneous part of the periodic grating, its nonlinearity is also calculated by the nonlinear GSM. For the calculation of the presented results $N_o = 21$ diffraction orders and $N_l = 138, \ldots, 275$ GSM-layers were used, each with height of $\approx 2\,\mathrm{nm}$.

Figure 7.5(a) shows the transmitted intensity at the FF for a TM polarized incident plane wave with wavelength λ_{FF}. The spectrum exhibits a smooth dependence on λ_{FF} and h_w, except for a narrow stripe that corresponds to a sharp decrease of the transmission. This change in transmission can clearly be attributed to the excitation of a TM_0 waveguide mode: the analytically (Eq. (2.43)) calculated resonance wavelength, which is represented by the pink line in the spectrum in Fig. 7.5(a), is in close proximity to the band of reduced transmission seen in the spectrum. The analytical calculation is obtained for a infinitesimally small perturbation of the slab waveguide, whence the agreement is nearly perfect for low h_w and slowly deteriorates as h_w, and therefore h_g, increases. Moreover, Fig. 7.5(d) shows the y-component of the magnetic field at the FF, H_y^{FF}, inside and around the textured waveguide, which is indicated by blue lines, corresponding to the black cross (1) in Fig. 7.5(a), i.e. to $\lambda_{\mathrm{FF}} = 3.861\,\mu\mathrm{m}$ and $h_w = 345\,\mathrm{nm}$. This field profile clearly matches that of the TM_0 mode. Note that for a 1D grating under non-conical incidence, the TE and TM components of the electromagnetic field decouple. Accordingly, there is only a single TM mode excited under TM-incidence, in the considered h_w-λ_{FF} domain.

To better visualize the SHG enhancement, Fig. 7.5(b) depicts the logarithm $\tilde{T}^{\mathrm{SH}} = \log_{10} T^{\mathrm{SH}}$ of the SH intensity, emitted into the substrate region, for the same incident fundamental field as in Fig. 7.5(a). The transmission coefficient, T^{SH}, is normalized to the SHG in a bulk layer of GaAs having the same volume as the textured grating. This fig-

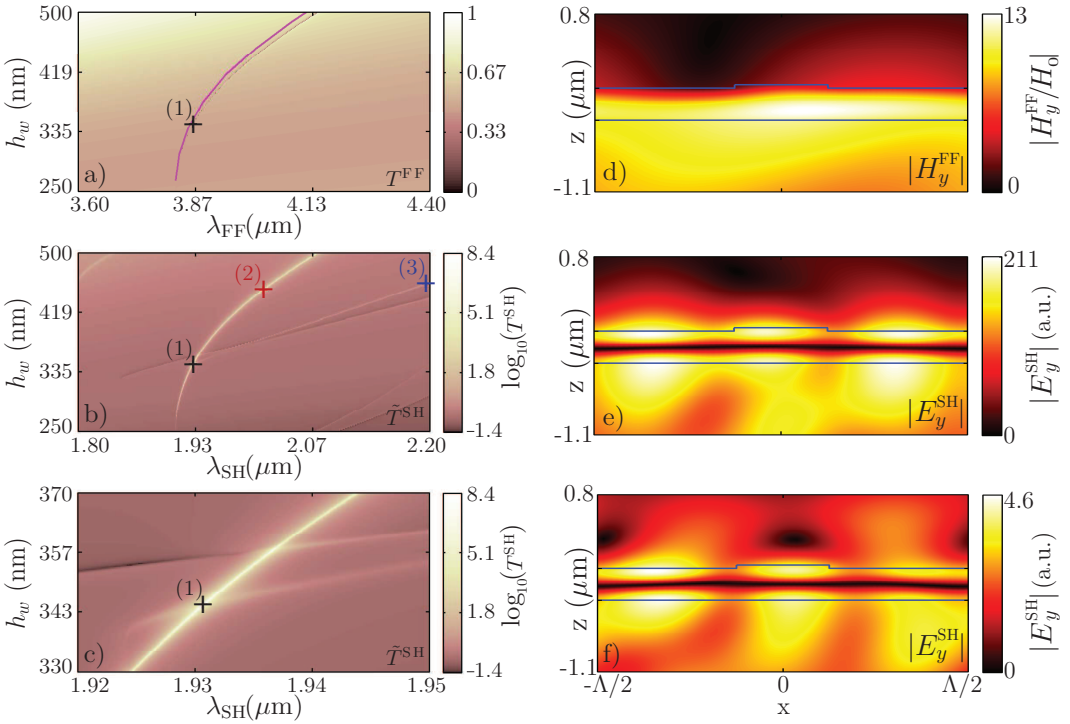

Figure 7.5: (a-c) Linear and nonlinear transmission spectra for 1D binary gratings. (d-f) Magnetic and electric field distributions at simultaneous resonance wavelength $\lambda_{\mathrm{FF}} = 3.45\,\mu\mathrm{m}$ for $h_w = 381\,\mathrm{nm}$; a) Spectra at the fundamental wavelength (pink dispersion curve indicates analytical calculation of TM_0 mode). b) Spectra at second harmonic wavelengths (logarithmic scale). c) Close-up of the region around a point where an intrinsic and inherited SH resonance dispersion curve cross. d) $|H_y^{\mathrm{FF}}|$ of the fundamental TM_0-resonance at the FF. e) $|E_y^{\mathrm{SH}}|$ of simultaneous resonance at SH. f) $|E_y^{\mathrm{SH}}|$ of intrinsic TE_1 mode at the SH wavelength $\lambda_{SH} = \lambda_{\mathrm{FF}}/2 = 1.725\,\mu\mathrm{m}$ for a TE polarized incident plane wave.

ure clearly shows an increase of the SH intensity due to the excitation of the inherited TM modes and intrinsic TE modes. No intrinsic TM modes are excited in this case. This is due to the particular symmetry properties of $\boldsymbol{\chi}^{(2)}$: for a TM polarized fundamental field $\mathbf{E}(\mathbf{r}) = (E_x(\mathbf{r}), 0, E_z(\mathbf{r}))$, only the y-component (i.e. the TE component) of the nonlinear polarization is non-vanishing.

At $\lambda_{SH} = 2.196\,\mu\mathrm{m}$ and $h_w = 460\,\mathrm{nm}$, corresponding to the blue cross (3) in Fig. 7.5(b), one can achieve an intensity enhancement of $T^{\mathrm{SH}} = 10^{0.55} \approx 3$ due to the excitation of an intrinsic TE_1 mode. Although the intensity is only three times higher than in the case of the bulk material, it is still significantly larger than the surrounding designs of textured waveguides. The excitation of the inherited mode yields much stronger enhancement of SHG, namely at $\lambda_{SH} = 2.011\,\mu\mathrm{m}$ and $h_w = 451\,\mathrm{nm}$ (red cross (2)), $T^{\mathrm{SH}} = 10^{4.78}$.

The strongest SH enhancement of $T^{\mathrm{SH}} = 10^{8.5}$ is achieved for simultaneous excitation of the inherited TM_0 mode and intrinsic TE_1 mode at $\lambda_{SH} = 1.931\,\mu\mathrm{m}$ and $h_w = 345\,\mathrm{nm}$ (black cross (1)). A close-up of the region of the doubly resonant excitation is presented

in Fig. 7.5(c). The electric field profile E_y^{SH} at the SH for this particular configuration, plotted in Fig. 7.5(e), shows high field enhancement inside the waveguide-grating region with two pronounced maxima at the top and bottom of the structure. This field profile can be explained by inspecting the distribution of the field E_y^{SH} of the intrinsic TE_1 mode at $\lambda_{SH} = 1.725\,\mu\mathrm{m}$ and $h_w = 381\,\mathrm{nm}$. Thus, for TE-polarized incident waves with $\lambda_{\mathrm{FF}} = 2\lambda_{SH} = 3.45\,\mu\mathrm{m}$, the electric field around the textured waveguide closely resembles the profile of a TE_1 mode, as can be seen in Fig. 7.5(f). However, the field enhancement is not as strong as in the case of the linear TM_0 mode, as per Fig. 7.5(b).

This analysis fully explains the dominant feature of the SH response of the 1D device when the two types of resonances are simultaneously excited, namely the overall SH intensity is mainly determined by the strong enhancement inside the waveguide of the fundamental field due to the linear resonance. The field profile, however, is determined by the profile of the intrinsic TE_1 mode.

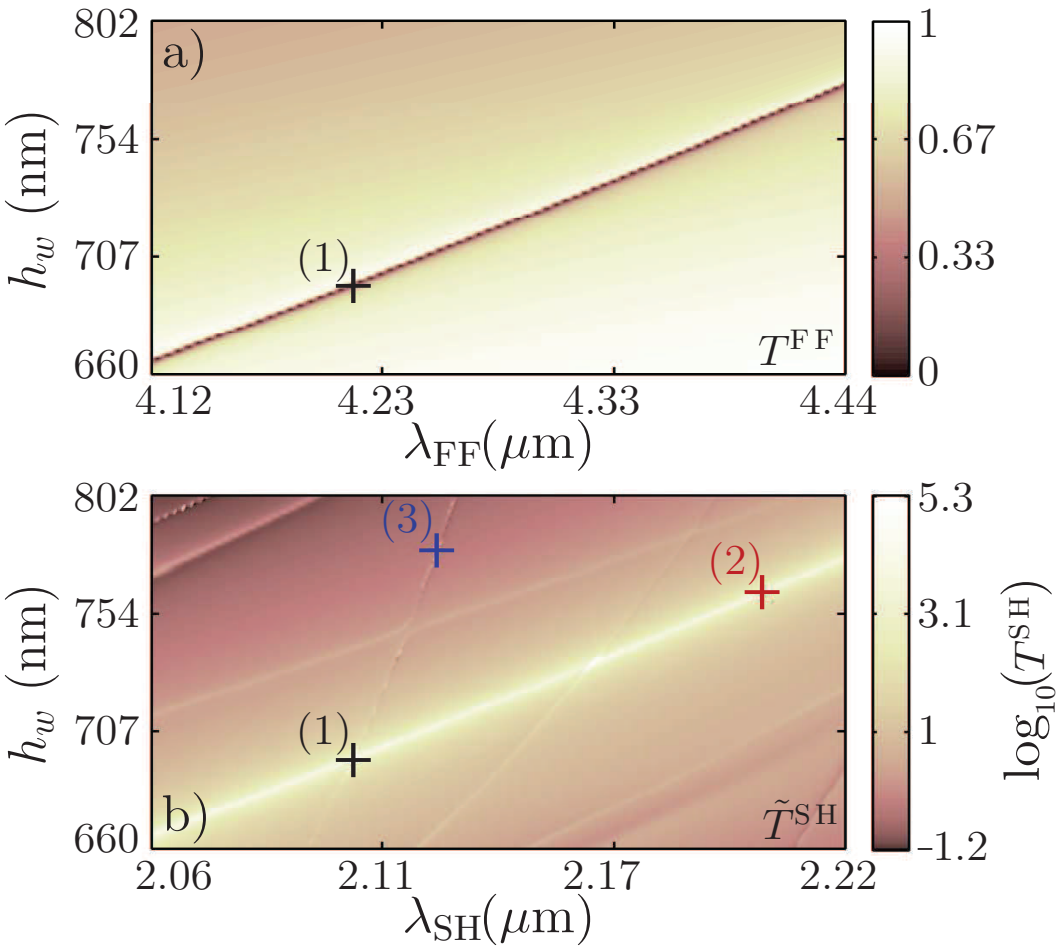

Figure 7.6: Spectra of transmission coefficient for the cylindrical texture determined for different waveguide heights under normal incidence. a) Spectra at the fundamental wavelength. b) Spectra at SH wavelengths (logarithmic scale).

7.5.2 Two-dimensional cylindrical gratings

The ideas presented in the previous section can be easily extended to a similar 2D device. To illustrate this, the cylindrical GaAs grating in Fig. 7.4(b) is considered. Its unit cell contains of a cylinder with radius, $r = 500\,\mathrm{nm}$ and relative height $h_g = 0.1h_w$, placed at the centre of a square with side-length $\Lambda = 2\,\mu\mathrm{m}$. The GaAs waveguide is placed on a CaF_2 substrate and its height, h_w, varies from $660\,\mathrm{nm}$ to $802\,\mathrm{nm}$. For the presented computationally expensive parameter sweep, a total of $N_o = 81$ diffraction orders and $N_l = 350, \ldots, 450$ GSM-layers with individual height of $2\,\mathrm{nm}$ were used. By checking the convergence with respect to N_o for a fixed $h_w = 700\,\mathrm{nm}$ and varying wavelength, it was found that $N_o = 81$ already yields qualitatively correct results.

The results for normally incident, x-polarized plane waves with wavelength $\lambda_{\mathrm{FF}} = 4.12\,\mu\mathrm{m}$ to $4.44\,\mu\mathrm{m}$ are summarized in Fig. 7.6. As in the 1D-case, the transmission spectrum at the FF reveals the resonant excitation of a TM_0 mode, seen in Fig. 7.6(a) as a dark stripe.

The spectrum of radiated SH in the direction of the incoming wave, presented in Fig. 7.6(b), exhibits distinct maxima along the dispersion curves of the inherited and intrinsic resonances. For example, for the intrinsic TE_0-resonance at $\lambda_{SH} = 2.126\,\mu\mathrm{m}$ and $h_w = 780\,\mathrm{nm}$ (blue cross (3)) the SHG enhancement as compared to a uniform GaAs layer with the same volume is approximately six-fold ($T^{\mathrm{SH}} = 10^{0.77}$). The inherited resonance at $\lambda_{SH} = 2.201\,\mu\mathrm{m}$ and $h_w = 763\,\mathrm{nm}$ (red cross (2)) yields $T^{\mathrm{SH}} = 10^{3.95}$.

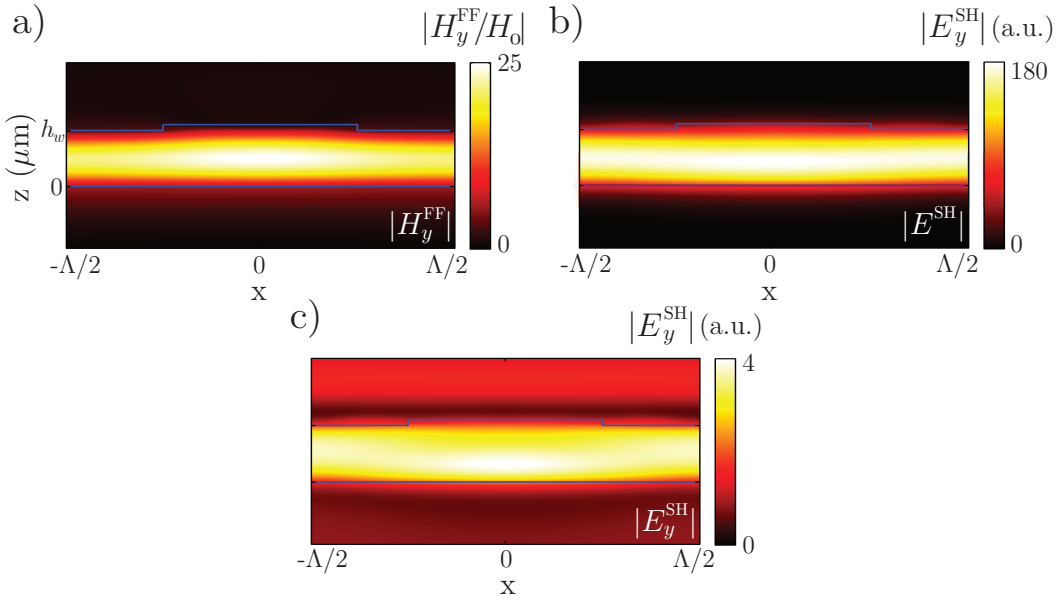

Figure 7.7: Magnetic and electric field distributions at simultaneous resonance wavelength $\lambda_{\mathrm{FF}} = 4.213\,\mu\mathrm{m}$ and $h_w = 696\,\mathrm{nm}$. a) $|H_y^{\mathrm{FF}}|$ of the fundamental TM_0-resonance at the FF. b) $|E_y^{\mathrm{SH}}|$ of simultaneous resonance at SH. c) $|E_y^{\mathrm{SH}}|$ of intrinsic TE_0 mode at SH wavelength $\lambda_{SH} = \lambda_{\mathrm{FF}}/2 = 2.107\,\mu\mathrm{m}$ for TM polarized incident plane wave.

This parameter study finds a 2D periodic device in which inherited and intrinsic resonances can be simultaneously excited, the corresponding parameters being $\lambda_{\mathrm{FF}} = 4.213\,\mu\mathrm{m}$ and $h_w = 696\,\mathrm{nm}$ (black cross (1)). This double resonance leads to an intensity enhancement of $I^{\mathrm{SH}} \approx 10^5$ as compared to the reference uniform layer. The spatial profile of the magnetic field, calculated in the x-z-plane passing through the centre of the cylinder ($y = 0$) is displayed in 7.7(a). The field distribution closely resembles the profile of the TM_0-mode, showing strong field enhancement and increased field confined in the waveguide as compared to the 1D-case. The simulations suggest that normal incidence and a larger waveguide thickness favour stronger confinement. Another aspect is the nonlinear electric field profile at the simultaneous resonance in Fig. 7.7(b), which is determined by two factors: Its magnitude is chiefly determined by the field enhancement at the FF. However, its shape and strong confinement are due to the excitation of the intrinsic nonlinear TE_0 mode Fig. 7.7(c), which is obtained for $\lambda_{SH} = 2.126\,\mu\mathrm{m}$ and $h_w = 780\,\mathrm{nm}$ (blue cross (3)).

7.6 Conclusions

This chapter presented an extension of the generalised source method to analyse SHG in 2D periodic structures containing non-centrosymmetric, quadratically nonlinear optical materials. The linear and nonlinear parts of the calculations are decoupled in the undepleted pump approximation and a three-step computational process was derived. The correct Fourier factorisation rule for inhomogeneous problems and its beneficial effect on the convergence in the nonlinear calculations was shown.

In a benchmark structure the asymptotic runtime estimate of $\mathcal{O}(N_0 \log(N_0) N_l \log(N_l))$ was validated experimentally, which renders the GSM especially suitable for modelling low permittivity contrast, dielectric gratings. The reason for the asymptotic runtime advantage of the GSM compared to the RCWA can be explained in two ways: Fundamentally, the GSM solves the grating problem for exactly one incident signal. The RCWA computes the $\mathcal{S}$ matrix for a grating structure, which is valid for any incident field, that can be described by linear combination of the incident diffraction orders Eq. (3.61). This is in most practical cases no necessary. From an algorithmic point of view, the GSM makes ample use of the Toeplitz-property of the system matrix A whence an efficient solution method in the form of GMRES can be employed to solve the grating problem.

As a practical application of the nonlinear GSM the design of 1D and 2D textured GaAs slab waveguides were optimized to achieve maximal radiated SH by simultaneous excitation of linear and nonlinear slab waveguide modes. In a addition to an analytical estimate for the resonance frequencies of the textured slab waveguide, investigation of the field profiles identified the nature of the resonances responsible for the remarkably high intensity enhancement of more than 8 orders of magnitude in comparison to the SH radiated by a uniform slab. Importantly, the formalism presented in this study can be extended to

several other key nonlinear optical processes, including surface SHG from centrosymmetric materials and third harmonic generation, which will be shown in the next two chapters.

Bibliography

[1] A. A. Shcherbakov and A. V. Tishchenko, "Fast numerical method for modelling one-dimensional diffraction gratings," IEEE J. Quantum Electron. **40**, 538 (2010).

[2] A. A. Shcherbakov and A. V. Tishchenko, "New fast and memory-sparing method for rigorous electromagnetic analysis of 2d periodic dielectric structures," J. Quant. Spectrosc. Radiat. Transfer **113**, 158–171 (2012).

[3] A. A. Shcherbakov and A. V. Tishchenko, "Efficient curvilinear coordinate method for grating diffraction simulation," Opt. Express **21**, 25236–25247 (2013).

[4] J. E. Sipe, "New green-function formalism for surface optics," J. Opt. Soc. Am. B. **4**, 481–489 (1987).

[5] A. Tishchenko, "A generalized source method for wave propagation," Pure and Applied Optics **7**, 1425 (1998).

[6] P. Lalanne and G. M. Morris, "Highly improved convergence of the coupled-wave method for tm polarization," J. Opt. Soc. Am. A. **13**, 779–784 (1996).

[7] L. Li, "Use of fourier series in the analysis of discontinuous periodic structures," J. Opt. Soc. Am. A. **13**, 1870–1876 (1996).

[8] L. Li, "New formulation of the fourier modal method for crossed surface-relief gratings," J. Opt. Soc. Am. A. **14**, 2758–2767 (1997).

[9] M. Weismann, D. F. Gallagher, and N. C. Panoiu, "Generalized source method for modeling nonlinear diffraction in planar periodic structures," in Spie photonics europe (International Society for Optics and Photonics, 2014), pp. 913115–913115.

[10] R. E. Blahut, *Fast Algorithms for Digital Signal Processing* (Addison-Wesley, 1985).

[11] Y. Saad and M. H. Schultz, "GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems," SIAM J. Sci. Stat. Comput. **7**, 856–869 (1986).

[12] MATLAB, *version 8.4 (R2014b)* (The MathWorks Inc., Natick, Massachusetts, USA, 2014).

[13] R. W. Boyd, *Nonlinear optics* (Academic Press, 2008).

[14] S. Bergfeld and W. Daum, "Second-Harmonic Generation in GaAs: Experiment versus Theoretical Predictions of $\chi^{(2)}_{xyz}$," Phys. Rev. Lett. **90**, 036801 (2003).

[15] N. C. Panoiu and R. M. Osgood, "Enhanced optical absorption for photovoltaics via excitation of waveguide and plasmon-polariton modes," Opt. Lett. **32**, 2825–2827 (2007).

[16] C. G. Biris and N. C. Panoiu, "Excitation of dark plasmonic cavity modes via nonlinearly induced dipoles: applications to near-infrared plasmonic sensing," Nanotechnol. **22**, 235502 (2011).

Chapter 8

Giant Nonlinear Optical Activity of Achiral Origin in Planar Metasurfaces with Quadratic and Cubic Nonlinearities

8.1 Introduction

In recent years, there has been a growing interest in the investigation of the nonlinear properties of plasmonic nanostructures and optical metamaterials, due to the associated strong enhancements of electromagnetic fields and to the capability to engineer the structural symmetry of their unit cell [1–4]. In particular, nonlinear optical circular dichroism of metamaterial and plasmonic nanostructures have been demonstrated [5–12]. However, the observed nonlinear optical CD in most experiments arises either from the chirality in 3D nanostructures, or from the extrinsic contribution in the case of 2D structures when the fundamental beam is incident at oblique angles. It should be pointed out that even seemingly 2D chiral materials, such as monolayers of chiral thin-film nanostructures are intrinsically 3D chiral as recently demonstrated in the linear [13] and in the nonlinear optical regimes [12]. In such nanostructures, chirality arises from the presence of a substrate on one side of the thin-films and not on the other.

In this chapter, giant circular dichroism for both second- and third-harmonic generation on ultrathin plasmonic metasurfaces with broken in-plane mirror symmetry is demonstrated [14]. Despite the symmetry breaking along the surface normal direction due to the presence of substrate, this dichroism is primarily of achiral origin. This result is verified by the simulation and measurement of harmonic generation for circularly polarized fundamental beam at normal incidence from both the front (air) and back (substrate) sides.

The remainder of this chapter derives the effective nonlinear susceptibilities for Trisceli- and Gammadion-type nanostructures in Section 8.2. Thereafter, Section 8.3 describes the experimental setup these structures and the computational approaches to modelling them are described. The simulation and measurements of the nonlinear CD is pre-

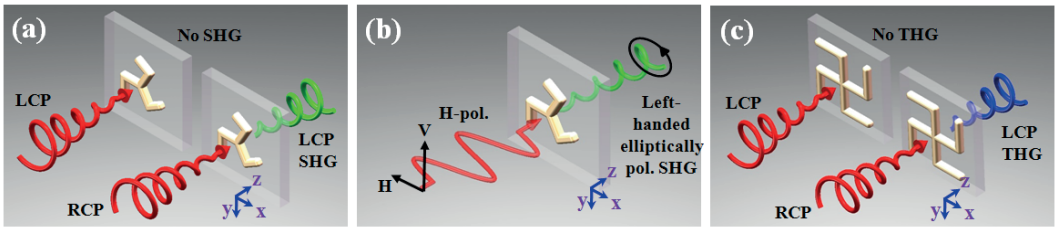

Figure 8.1: Schematic of nonlinear optical dichroism from Trisceli- and Gammadion-type plasmonic nanostructures with three- ($\mathscr{C}_3$) and fourfold ($\mathscr{C}_4$) rotational symmetries. (a) and (c) illustrate the nonlinear circular dichroism for illumination with LCP and RCP polarizations on such structures which will lead to SHG and THG with opposite polarizations but only for the RCP illumination. Trisceli-type nanostructures with near unity SHG-CD shown in (b) would generate left-handed elliptically polarized SHG light when illuminated by linear polarized light.

.

sented in Section 8.4 before final conclusions are drawn in Section 8.5.

8.2 Theory of effective nonlinear susceptibility for Trisceli- and Gammadion-type nanostructures

The strong nonlinear CD of achiral origin is explained by the effective nonlinear susceptibility coefficients associated with the specific $\mathscr{C}_3$ and $\mathscr{C}_4$ rotational symmetries of Trisceli- and Gammadion-type plasmonic nanostructures, which only involve the in-plane tensor elements As shown schematically in Fig. 8.1, the magnitude of the measured nonlinear CD for second- and for third-harmonic generation reaches near the maximum theoretical limit. Additionally, we show that upon illumination with left- or right-handed circularly polarized light, the $\mathscr{C}_3$ and $\mathscr{C}_4$ rotational symmetric nanostructures can generate cross-polarized SHG and THG waves, respectively.

To maintain the rotational symmetry of the overall experimental configuration, we limit ourselves to the configuration in which the fundamental wave is incident normally onto the surface. Therefore, the relevant effective nonlinear susceptibility tensor elements are those with sub-indices consisting only of x and y, assuming a Cartesian coordinate system, where the sample is in the x-y plane. Due to the rotational symmetry of both the individual structure and the lattice, the effective nonlinear susceptibility tensor can be reduced to just a few independent elements.

8.2.1 Effective second order susceptibility for $\mathscr{C}_3$ rotational symmetry

We start with the second-order nonlinear susceptibility tensor for the metasurfaces consisting of nanostructures with $\mathscr{C}_3$ rotational symmetry. It has been shown [15] that the $\mathscr{C}_3$ rotational symmetry reduces the tensor elements to two independent values:

$$\chi^{(2)}_{yyy} = -\chi^{(2)}_{yxx} = -\chi^{(2)}_{xxy} = -\chi^{(2)}_{xyx} = \chi^{(2)}_{1}, \tag{8.1a}$$

$$\chi^{(2)}_{xxx} = -\chi^{(2)}_{xyy} = -\chi^{(2)}_{yyx} = -\chi^{(2)}_{yxy} = \chi^{(2)}_{2}. \tag{8.1b}$$

In contrast, for previously investigated crystal surfaces possessing $\mathscr{C}_{3v}$ symmetry, such as Si (111) and GaAs (111), both the rotational symmetry and mirror symmetry are present and only a single tensor element is nonzero depending on whether x or y is aligned with one of the symmetry planes. Thus, for the $\mathscr{C}_3$ structure investigated here, the effective nonlinear polarization can be expressed as:

$$P^{2\omega}_{x} = \epsilon_0\chi_2\left(E^2_x - E^2_y\right) - 2\epsilon_0\chi_1 E_x E_y \tag{8.2a}$$

$$P^{2\omega}_{y} = -2\epsilon_0\chi_2 E_x E_y - \epsilon_0\chi_1\left(E^2_x - E^2_y\right) \tag{8.2b}$$

For a circularly polarized incident beam, the electric field vector is given by $\mathbf{E} = E_0\left(\mathbf{e}_x \pm i\mathbf{e}_y\right)/\sqrt{2}$, where the $+$ and $-$ sign correspond to LCP and RCP, respectively. Here we use the definition of LCP and RCP as defined from the point of view of the receiver. The effective nonlinear polarization can be expressed as

$$\mathbf{P}^{2\omega} = \epsilon_0\left(\chi_2 \mp i\chi_1\right)\left(\mathbf{e}_x \pm i\mathbf{e}_y\right)E^2_0. \tag{8.3}$$

This equation satisfies the selection rule studied previously, circularly polarized light cannot produce the SHG with the same polarization as the fundamental light at normal incidence onto a structure with $\mathscr{C}_3$ symmetry [16–20]. In addition, Eq. (8.3) shows that the intensities of the SHG are generally different for the two circular polarization states. However, to obtain a difference in the SHG intensity for the two circular polarization states it is necessary that χ_1 and χ_2 are different in phase. Note that if χ_1 and χ_2 have the same phase, a coordinate rotation can be employed to reduce the in-plane nonlinear coefficients to a single one $\chi = \sqrt{\chi^2_1 + \chi^2_2}$ and χ_2, i e., the system becomes the same as for a $\mathscr{C}_{3v}$ symmetry. The highest contrast between the two circular polarization states can be achieved in the case that χ_1 and χ_2 are equal in amplitude and $\pi/2$ out of phase. This would lead to a unity SHG-CD, which arises from the cancellation of the nonlinear polarization between the contributions from χ_1 and χ_2 for one circular polarization state, but a constructive interference between their contributions for the other circular polarization. The above analysis shows that, for a nanostructured thin film with $\mathscr{C}_3$ rotational symmetry, nonlinear susceptibility tensor components with z indices are not necessary for introducing nonlinear CD for second-harmonic generations. This is in stark contrast to commonly studied SHG-CD effect in chiral thin films [21].

For an incident wave with linear polarization along the horizontal direction $(E_x, E_y) = (E_0, 0)$, we have $P^{2\omega}_x = \epsilon_0\chi_2 E^2_0$ and $P^{2\omega}_y = -\epsilon_0\chi_1 E^2_0$. If an analyser is placed in trans-

mission with orientation angle , the measured SHG intensity is given by:

$$I_\theta^{2\omega} \propto |P_\theta^{2\omega}|^2 \propto |\chi_2 \cos\theta - \chi_1 \sin\theta|^2 \tag{8.4}$$

By normalizing χ_1 to unity and setting $\chi_2 = a\exp(i\varphi)$, the above equation can be rewritten in a simple sinusoidal form as:

$$I_\theta^{2\omega} \propto 1 - \cos\beta \sin(2\theta + \gamma) \tag{8.5}$$

where

$$\beta = \sin^- 1\left(\frac{2a\sin\varphi}{1+a^2}\right) \text{ and } \gamma = \tan^- 1\left(\frac{a^2-1}{2a\cos\varphi}\right). \tag{8.6}$$

Interestingly, based on Eq. (8.3) and the definition of β, SHG-CD can be derived as:

$$\mathrm{CD} = \frac{I_{\mathrm{LCP}}^{2\omega} - I_{\mathrm{RCP}}^{2\omega}}{I_{\mathrm{LCP}}^{2\omega} + I_{\mathrm{RCP}}^{2\omega}} = \sin\beta \tag{8.7}$$

Thus, Eq. (8.5) and Eq. (8.7) show that the circular dichroism for SHG can be independently obtained by two measurements - the ellipticity of the SHG for a linearly polarized fundamental beam, or a direct measurement of the relative difference between the SHG signals of the two opposite circularly polarized incident fundamental beams.

8.2.2 Effective third order susceptibility for $\mathscr{C}_4$ rotational symmetry

For THG processes, applying a $\mathscr{C}_4$ rotational symmetry reduces the third-order nonlinear tensors to four independent elements. Hence, the third-order nonlinear polarization can be expressed as:

$$P_x^{3\omega} = \epsilon_0\left[\chi_1 \mathbf{E}_x^3 + \chi_2 E_x E_y^2 + \chi_3 E_x^2 E_y + \chi_4 E_y^3\right] \tag{8.8a}$$

$$P_y^{3\omega} = \epsilon_0\left[\chi_1 \mathbf{E}_y^3 + \chi_2 E_y E_x^2 - \chi_3 E_y^2 E_x - \chi_4 E_x^3\right] \tag{8.8b}$$

where [15]

$$\chi_1 = \chi_{xxxx}^{(3)}, \tag{8.9a}$$

$$\chi_2 = \chi_{xxyy}^{(3)} + \chi_{xyxy}^{(3)} + \chi_{xyyx}^{(3)}, \tag{8.9b}$$

$$\chi_3 = \chi_{xxxy}^{(3)} + \chi_{xxyx}^{(3)} + \chi_{xyxx}^{(3)}, \tag{8.9c}$$

$$\chi_4 = \chi_{xyyy}^{(3)}. \tag{8.9d}$$

8.3 Simulation and experimental setup

The section briefly overviews the specification and fabrication of the devices, the experimental measurement process and the simulation methods used.

8.3.1 Experimental setup

The Trisceli- and Gammadion-type plasmonic nanostructures which have $\mathscr{C}_3$ and $\mathscr{C}_4$ symmetries, respectively, are fabricated on ITO coated BK7 glass substrates by using electron-beam lithography and a lift-off process. As shown in Fig. 8.2, these two types of gold nanostructures with thickness of $30\,\mathrm{nm}$ are arranged in triangular and square lattices with periods of 400 and $500\,\mathrm{nm}$, respectively. The Gammadion nanostructures are subsequently covered by a thin layer of an organic conjugated polymer (PFO) with high third-order optical nonlinearity to enhance the efficiency of THG.

In most conventional SHG-CD measurements, oblique incidence of light is usually employed to project a strong electric field component of light along the z-direction of the sample. This geometry allows access to nonlinear susceptibility tensor components with one or several z indices, which can be quite large in some materials [22, 23]. In particular, the so-called chiral tensor components contain such z indices. Since we seek to minimize the chiral contribution to SHG-CD and to emphasize an achiral SHG-CD instead, here, we use a femtosecond laser beam under normal incidence to the sample surface to interact with the nanostructures. Moreover, we used a low numerical aperture objective ($\mathrm{NA} = 0.1$) to minimize the projection of a polarization component along the propagation direction. The signals of harmonic generation are then collected in the transmission direction. All linear and nonlinear measurements have been carried out by S. Chen, F. Zeuner and G. Li at the University of Birmingham and subsequently at the University of Paderborn.

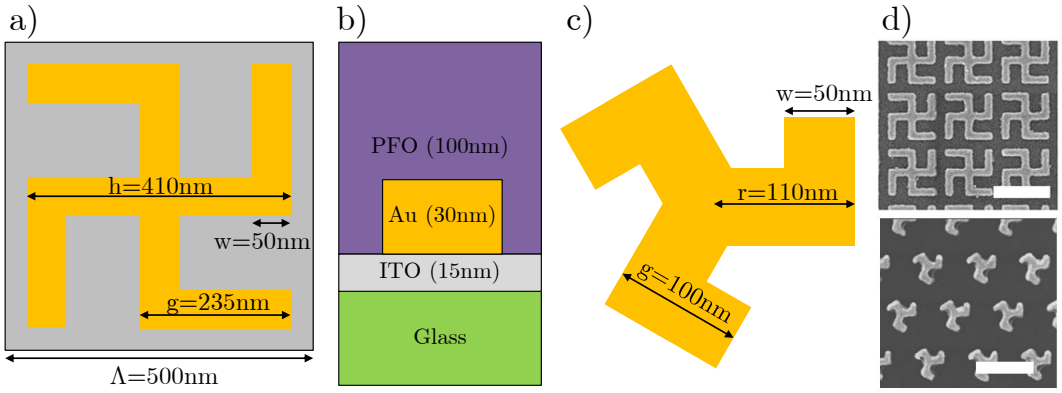

Figure 8.2: Geometric parameters of Gammadion- (a,b) and Trisceli-type structures (c). The z-extend of the Trisceli-type structures is as that of the Gammadion-type structure, but it are not covered with PFO. The white bars and the scanning electron microscopy (SEM) images (d) indicate $500\,\mathrm{nm}$. [Samples have been fabricated and characterised by SEM imaging by G. Li at the University of Birmingham.]

8.3.2 Simulation setup

The nonlinear optical response of the metasurfaces at the TH wavelength has been numerically calculated using the generalised source method described.

A few modifications to its formulations in Chapter 7 have been made: The three step solution process of the undepleted pump approximation from Section 7.2 is maintained, however the calculations at the fundamental frequency were performed by means of the RCWA. This is to avoid the large runtime of GSM for metallic structures at longer wavelengths. The second modification is the incorporation of the TH source polarization

$$\mathbf{P}^{3\omega}(\mathbf{r}) = \epsilon_0 \chi^{(3)}(\mathbf{r}) \left(\mathbf{E}(\mathbf{r}) \cdot \mathbf{E}(\mathbf{r})\right) \mathbf{E}(\mathbf{r}) \tag{8.10}$$

in step two of the calculation instead of Eq. (7.2). The nonlinear source polarisation $\mathbf{P}^{3\omega}(\mathbf{r})$ depends on the complex field $\mathbf{E}(\mathbf{r})$ and the nonlinear susceptibility $\chi^{(3)}$ of the respective material at the location $\mathbf{r}$. In particular, it is assumed that the optical nonlinearity of the medium is homogeneous and isotropic, so that the nonlinear third-order susceptibility tensor is described by a single quantity. The numerical values of this susceptibility in the cases of Gold [24], PFO [25], and ITO [26] are

$$\chi^{(3)}_{\mathrm{Gold}} = 5.58 \times 10^{-20}\,\mathrm{m\,V^{-2}}, \tag{8.11a}$$

$$\chi^{(3)}_{\mathrm{PFO}} = \chi^{(3)}_{\mathrm{Gold}}, \text{ and} \tag{8.11b}$$

$$\chi^{(3)}_{\mathrm{ITO}} = 10^{-20}\mathrm{m\,V^{-2}}, \tag{8.11c}$$

respectively. The linear material dispersion parameters are taken from Refs. [27, 28]

All simulation results for the Gammadion-type structure presented hereafter have been obtained by the GSM (and RCWA at the fundamental wavelength) with a maximum of $N_1 = N_2 = 17$ harmonics and $N_L = 100$ GSM-layers for which convergence was reached.

In the case of SHG by the Trisceli-type structures, the convergence of GSM is very slow due to the fact that the nonlinear effects come from the interfaces between gold and the surrounding dielectric media (air and glass) and the high losses of Gold in the respective SH wavelengths. Instead, the SHG calculation is based on linear simulations at the fundamental and second harmonic wavelengths by using COMSOL Multiphysics [29] following the method proposed by O'Brien et al. [30, 31], in which the SHG is calculated by an overlapping integral between the nonlinear polarization at the metal-dielectric interfaces and the near field distribution of the electric field when the metasurface is illuminated by a plane wave at the second harmonic frequency. Simulations of the Trisceli-type structure have been performed by S. Chen at the University of Birmingham.

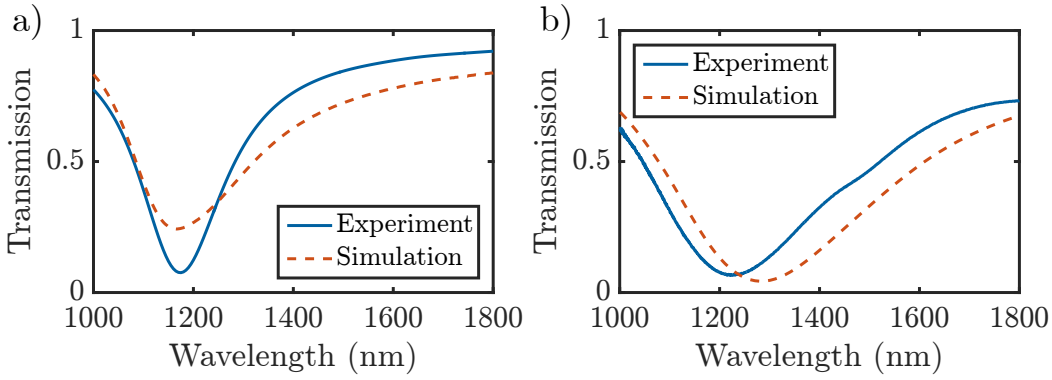

Figure 8.3: Measured ("Experiment") linear transmission spectra of the Trisceli- and PFO-coated Gammadion-type nanostructures for linearly (horizontally) polarized light, exhibiting localized plasmon resonance at wavelength of 1165 and 1230 nm, respectively, which agree well with the calculated (Simulation) transmission spectra. [Measurements of linear spectra have been performed by S. Chen, F. Zeuner and G. Li at the University of Birmingham. Simulations of the Trisceli-type structure have been performed by G. Li.]

8.4 Experimental and numerical investigation of Gammadion- and Trisceli-type structures

8.4.1 Linear and nonlinear optical results

Figure 8.3 shows the measured transmission spectra of the plasmonic nanostructures for horizontally polarized incident and transmitted light (HH) by using Fourier transform infrared spectrometer. The dips in the transmission spectra correspond to the excitation of localized surface plasmon polariton modes of the Trisceli- and Gammadion-type plasmonic nanostructures at wavelength of 1165 nm and 1230 nm, respectively, which are also confirmed by the numerical simulations: The calculated transmission spectrum for the Trisceli-type structures agree well with the experiment in the location of the resonance, but underestimates its spectral width. The calculated plasmon resonance for the Gammadion-type structure exhibits a similar spectral width. However, it is slightly shifted towards higher wavelengths.

To characterize the frequency dependent nonlinear optical activity, the wavelength of illuminating laser is tuned between $\lambda_0 = 1100\,\mathrm{nm}$ and $\lambda_0 = 1400\,\mathrm{nm}$ and the circularly polarized SHG and THG signals are recorded with a spectrometer Fig. 8.4(a,d).

We found that SHG and THG signals from the plasmonic nanostructures have opposite polarization states as compared to that of the illuminating laser. However, we also measured extremely low SHG and THG with the same polarization state as that of the illuminating laser although they are theoretically forbidden in the dipole approximation but can be observed due to the imperfections of nanofabrication [10, 11].

For the SHG measurements, both the L-R and R-L spectra increase with wavelength up

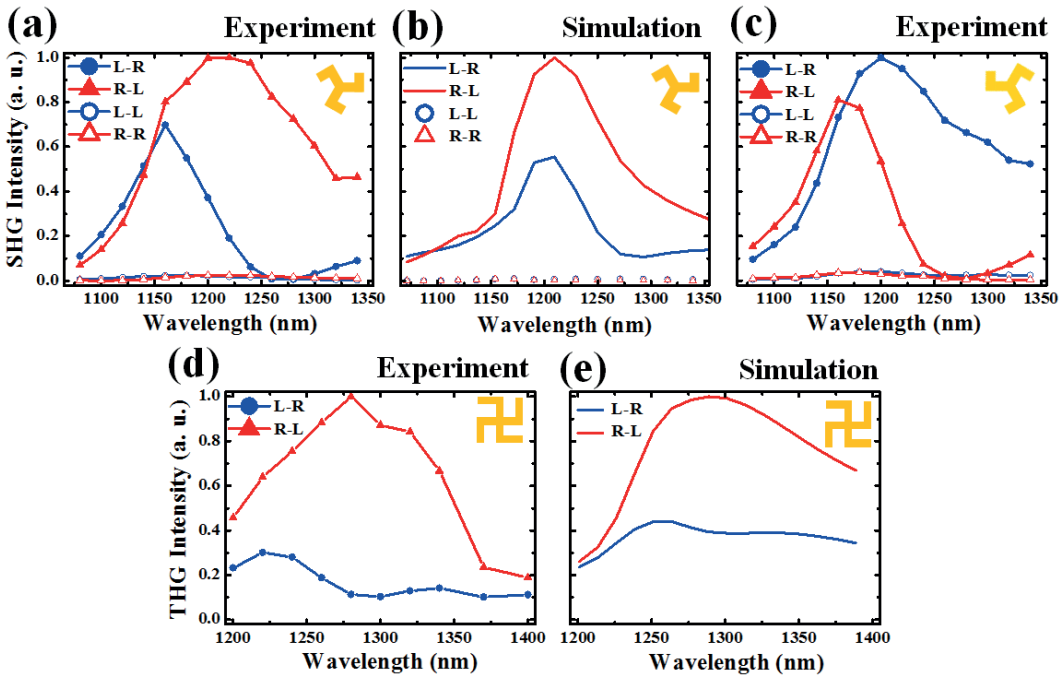

Figure 8.4: Polarization-dependent SHG and THG spectra of the Trisceli- and Gammadion-type plasmonic nanostructures. Measured SHG and THG spectra with same (LCP-LCP and RCP-RCP) and opposite circular polarizations (LCP-RCP and RCP-LCP) as that of the illuminating light. Note, in the figure we used the abbreviation of L for LCP and R for RCP, respectively. a,b) Measured (a) and numerically simulated (b) SHG spectra for fundamental wave normally incident from Trisceli-type nanostructure direction to substrate direction (front illumination). c) Measured SHG spectra for light incident from the substrate side (back illumination), which are very close to the results of front illumination but swapped between R-L and L-R. d,e) Measured (d) and simulated (e) THG spectra under front illumination show pronounced difference between R-L and L-R for fundamental wavelength with a bandwidth of around 150 nm. [Measurements of nonlinear spectra have been performed by S. Chen, F. Zeuner and G. Li at the University of Birmingham. Simulations of the Trisceli-type structure have been performed by S. Chen.]

to 1160 nm. However, the L-R intensity drops abruptly with further increasing wavelength above 1160 nm, while the R-L intensity keeps further increasing and peaks around 1220 nm. As a result, there exists a large difference between the SHG intensities for the two circularly polarized incident beams for wavelength above 1160 nm. Particularly, at around 1280 nm wavelength, the SHG for the illumination with LCP light (L-R) drops almost to zero while that for the illumination with RCP light (R-L) is very strong, leading to near unity value of 0.98 for SHG-CD at this wavelength.

Our measurements nicely confirm numerical simulations of the SHG and THG signals from such structures, which also predict large differences between the two polarized signals Fig. 8.4(b,e). The measured nonlinear spectra of the SHG intensity show similar features as those of calculated SHG spectra. The slight discrepancy between the experiment and the simulation might be due to both the deviation of the fabricated sample from the ideal

design, and the limited precision of the simulation of the nonlinear signals due to their extreme sensitivity to the local field distribution.

In an ideal achiral configuration of free standing $\mathscr{C}_3$ nanostructures without the presence of substrate, it is expected that when the sample is flipped between the front and reverse-side, SHG spectra are swapped between the two circular polarizations, i.e., $I^{\mathrm{f}}_{\mathrm{R}\to\mathrm{L}}(2\omega) = I^{\mathrm{b}}_{\mathrm{L}\to\mathrm{R}}(2\omega)$ and $I^{\mathrm{f}}_{\mathrm{L}\to\mathrm{R}}(2\omega) = I^{\mathrm{b}}_{\mathrm{R}\to\mathrm{L}}(2\omega)$ where "f" and "b" correspond to front- and back-illumination of the fundamental wave, respectively. This is because flipping the sample leads to a change of sign for either χ_1 or χ_2 (depending on flipping the sample by rotating along x or y axis), but not both, and consequently results in reversal of the CD effect based on Eq. (8.3). However, in the realistic case, the nanostructures are supported by a substrate which breaks the symmetry along the z direction. To investigate into the role of the symmetry breaking in the z-direction due to the presence of substrate on SHG-CD, we perform an additional SHG measurement on $\mathscr{C}_3$ sample, now flipped such that the fundamental beam is incident from the substrate side, with the spectra shown in Fig. 8.4(c). As can be seen from the figure, flipping the sample leads to exchange of the SHG spectra between the two circularly polarized fundamental waves. This observation serves as direct evidence that the observed SHG-CD is mainly due to the in-plane symmetry breaking alone, which is of achiral nature.

For the $\mathscr{C}_4$ sample, the THG spectra also show pronounced difference between the two circular polarization states over a broad frequency range between $1200\,\mathrm{nm}$ and $1400\,\mathrm{nm}$ Fig. 8.4(d). While the THG spectrum for the R-L configuration exhibits a peak around $1280\,\mathrm{nm}$ nm wavelength, the L-R spectrum is almost featureless over the entire frequency range. The THG intensity of the R-L configuration is approximately one order of magnitude greater than that of the L-R configuration at $1280\,\mathrm{nm}$. Again, the simulation of the THG signals exhibits similar features to those of the experimental results. However, the relative differences between the THG of the two circular polarizations are not as large as the measured values.

Based on the measured spectral results for both SHG and THG, as shown in Fig. 8.5(a,c), we calculated the nonlinear CDs by $(I_{\mathrm{LCP}} - I_{\mathrm{RCP}})/(I_{\mathrm{LCP}} + I_{\mathrm{LCP}})$. The SHG-CD varies from negative to positive values, crossing zero around $1150\,\mathrm{nm}$ nm and reaching near-unity value of 0.98 at $1280\,\mathrm{nm}$. On the other hand, the THG-CD reaches its maximum value of 0.79 at $1280\,\mathrm{nm}$. We note that this is the first observation of THG-CD on planar chiral plasmonic structures. Furthermore, the maximum value of the measured nonlinear SHG-CD is much higher than that of the linear CD effect observed previously from the Gammadion-type plasmonic nanostructures [13].

We also found very similar trends of the nonlinear CDs from the simulated SHG and THG signals Fig. 8.5(b,d), where the simulated maximum values of the SHD-CD and the

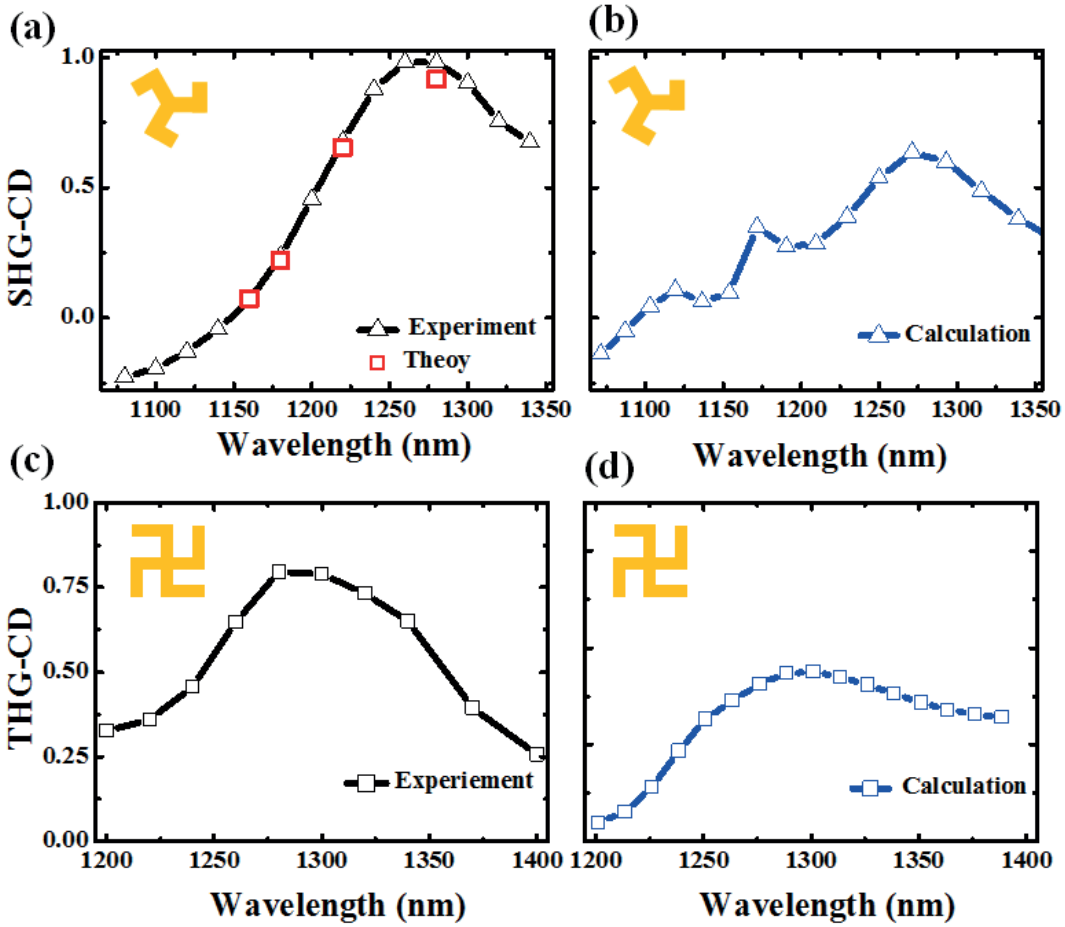

Figure 8.5: SHG-CD and THG-CD from the Trisceli- and Gammadion-type plasmonic nanostructures. a,c) Measured (Experiment) wavelength-dependent SHG-CD and THG-CD. The values for the squares in (a) are obtained from the calculation of SHG-CD based on the experimentally determined values of the effective nonlinear susceptibilities. b,d) Corresponding numerically simulated (Calculation) nonlinear CD spectra for the structures. Both of the calculated SHG and THG spectra show similar trend as the measured ones. While the peak of SHG-CD is close to the measured value, the calculated THG has a much higher deviation from our experimental data, which can be attributed to the imperfections of nanofabrication. [Measurements of nonlinear spectra have been performed by S. Chen, F. Zeuner and G. Li at the University of Birmingham. Simulations of Trisceli-type structures have been performed by G. Li.]

THG-CD are 0.637 at 1271 nm and 0.435 at 1301 nm, respectively. These values are somewhat lower than their respective measured values.

Above we showed that the nonlinear CD has to be related to the nonlinear susceptibility tensor elements. In order to retrieve the effective susceptibilities of the SHG from the Trisceli-type nanostructures, the SHG intensity as function of the rotational angle of the analyser is measured for four different wavelengths: 1160 nm, 1180 nm, 1220 nm, and 1280 nm, for a fixed linear polarization of the illuminating light along the vertical direction. As illustrated in Fig. 8.6, the SHG intensities show a sinusoidal dependence on the

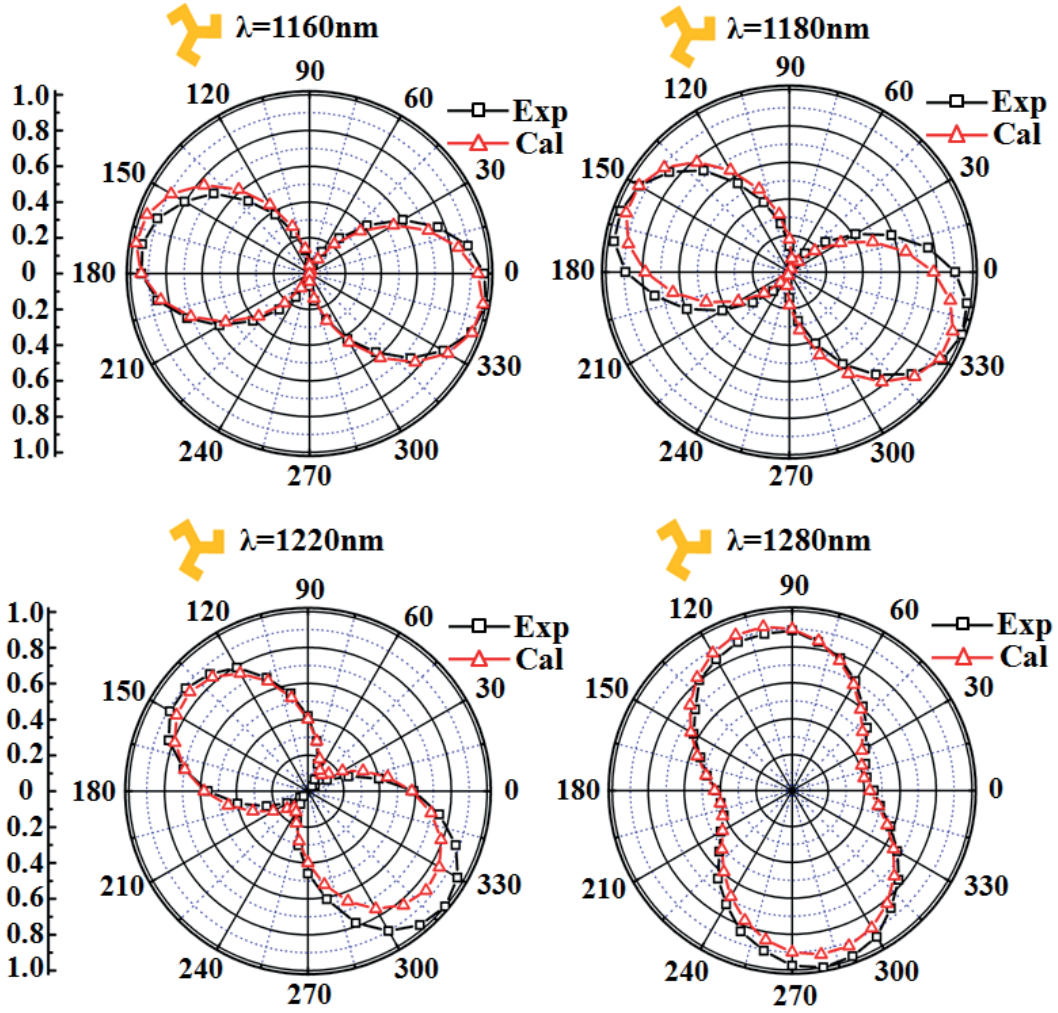

Figure 8.6: Intensity of the SHG signal as function of the rotation angle of the analyser for Trisceli-type plasmonic nanostructures. The results are obtained for linearly vertically polarized illumination at wavelengths of 1160, 1180, 1220, and 1280 nm, respectively. The square and triangular symbols represent the measured ("Exp") and calculated ("Cal") results.[Nonlinear measurements have been performed by S. Chen, F. Zeuner, and G. Li at the University of Birmingham. Simulations of Trisceli-type structures have been performed by G. Li.]

analyser angle. At the shortest wavelength of 1160 nm, the minimum value is close to zero, indicating that the SHG signal is linearly polarized at this wavelength. With the increase of wavelength, the minimum value rises, indicating the increase of the SHG ellipticity. This is consistent with the observation of vanishing SHG-CD at 1150 nm and its increase up to 1280 nm, as shown in Fig. 8.5.

By using Eq. (8.2), we can fit the curves and retrieve the values of χ_2 with χ_1 being normalized to 1. This normalization of χ_1 is well justified since we are only interested in the structural effect but not the absolute nonlinear signal strength, and consequently only the ratio between χ_2 with χ_1 is of relevance. From this retrieval we obtain for χ_2: $4.5 + 0.8i$,

$2+0.6i$, $0.9+0.8i$, and $0.13+0.68i$ for the fundamental waves at wavelength of $1160\,\mathrm{nm}$, $1180\,\mathrm{nm}$, $1220\,\mathrm{nm}$, and $1280\,\mathrm{nm}$, respectively. Based on Eq. (8.3) and the fitted values of χ_1 and χ_2, we numerically calculate the SHG-CD (red squares in Fig. 8.5(a)), which agrees very well with the directly measured SHG-CD. Thus, with two independent measurements, we confirm the presence of near-unity SHG-CD at wavelength of $1280\,\mathrm{nm}$ for planar chiral nanostructures.

8.5 Conclusions

In this chapter, near-unity nonlinear circular dichroism for both second- and third-harmonic generation from suitably designed ultrathin Trisceli- and Gammadion-type plasmonic nanostructures with three- and four-fold rotational symmetry we demonstrate. These two kinds of nanostructures allow a symmetry controlled generation of circularly polarized SHG and THG, respectively. Both giant SHG-CD and THG-CD are experimentally observed for the planar plasmonic nanostructures with negligible linear optical activity. These experimental results have been confirmed by numerical simulations using - in part - the numerical methods developed in this work. Importantly, the observed SHG-CD is identified to have an achiral origin. The simple fabrication of 2D plasmonic structures and wider applicability of SHG and THG techniques also enable more freedoms in designing chip-type nonlinear optoelectronic devices.

Bibliography

[1] W. Cai, A. P. Vasudev, and M. L. Brongersma, "Electrically Controlled Nonlinear Generation of Light with Plasmonics," Science **333**, 1720 (2011).

[2] M. Kauranen and A. V. Zayats, "Nonlinear plasmonics," Nat. Photon. **6**, 737–748 (2012).

[3] M. Hentschel, T. Utikal, H. Giessen, and M. Lippitz, "Quantitative Modeling of the Third Harmonic Emission Spectrum of Plasmonic Nanoantennas," Nano Lett. **12**, 3778–3782 (2012).

[4] M. Lapine, I. V. Shadrivov, and Y. S. Kivshar, "Colloquium: Nonlinear metamaterials," Reviews of Modern Physics **86**, 1093–1123 (2014).

[5] V. K. Valev, A. V. Silhanek, N. Verellen, W. Gillijns, P. Van Dorpe, O. A. Aktsipetrov, G. A. E. Vandenbosch, V. V. Moshchalkov, and T. Verbiest, "Asymmetric Optical Second-Harmonic Generation from Chiral G-Shaped Gold Nanostructures," Phys. Rev. Lett. **104**, 127401 (2010).

[6] A. Belardini, M. C. Larciprete, M. Centini, E. Fazio, C. Sibilia, D. Chiappe, C. Martella, A. Toma, M. Giordano, and F. Buatier de Mongeot, "Circular Dichroism in the Optical Second-Harmonic Emission of Curved Gold Metal Nanowires," Phys. Rev. Lett. **107**, 257401, 257401 (2011).

[7] M. J. Huttunen, G. Bautista, M. Decker, S. Linden, M. Wegener, and M. Kauranen, "Nonlinear chiral imaging of subwavelength-sized twisted-cross gold nanodimers," Opt. Mater. Express **1**, 46–56 (2011).

[8] A. Rose, D. A. Powell, I. V. Shadrivov, D. R. Smith, and Y. S. Kivshar, "Circular dichroism of four-wave mixing in nonlinear metamaterials," Phys. Rev. B **88**, 195148, 195148 (2013).

[9] S. P. Rodrigues, S. Lan, L. Kang, Y. Cui, and W. Cai, "Nonlinear Imaging and Spectroscopy of Chiral Metamaterials," Adv. Mater. **26**, 6157+ (2014).

[10] B. Metzger, L. Gui, J. Fuchs, D. Floess, M. Hentschel, and H. Giessen, "Strong Enhancement of Second Harmonic Emission by Plasmonic Resonances at the Second Harmonic Wavelength," Nano Lett. **15**, 3917–3922 (2015).

[11] V. K. Valev, J. J. Baumberg, B. De Clercq, N. Braz, X. Zheng, E. J. Osley, S. Vandendriessche, M. Hojeij, C. Blejean, J. Mertens, C. G. Biris, V. Volskiy, M. Ameloot, Y. Ekinci, G. A. E. Vandenbosch, P. A. Warburton, V. V. Moshchalkov, N. C. Panoiu, and T. Verbiest, "Nonlinear superchiral meta-surfaces: tuning chirality and disentangling non-reciprocity at the nanoscale," Adv. Mater. **26**, 4074–4081 (2014).

[12] R. Kolkowski, L. Petti, M. Rippa, C. Lafargue, and J. Zyss, "Octupolar Plasmonic Meta-Molecules for Nonlinear Chiral Watermarking at Subwavelength Scale," ACS Photonics **2**, 899–906 (2015).

[13] M. Kuwata-Gonokami, N. Saito, Y. Ino, M. Kauranen, K. Jefimovs, T. Vallius, J. Turunen, and Y. Svirko, "Giant Optical Activity in Quasi-Two-Dimensional Planar Nanostructures," Phys. Rev. Lett. **95**, 227401, 227401 (2005).

[14] S. Chen, F. Zeuner, M. Weismann, B. Reineke, G. Li, V. K. Valev, K. W. Cheah, N. C. Panoiu, T. Zentgraf, and S. Zhang, "Giant Nonlinear Optical Activity of Achiral Origin in Planar Metasurfaces with Quadratic and Cubic Nonlinearities," Adv. Mater. (2016).

[15] R. W. Boyd, *Nonlinear optics* (Academic Press, 2008).

[16] S. Bhagavantam and P. Chandrasekhar, "Harmonic generation and selection rules in nonlinear optics," in Proc. Indian Acad. Sci. A, Vol. 76, 1 (Springer, 1972), pp. 13–20.

[17] O. E. Alon, V. Averbukh, and N. Moiseyev, "Selection Rules for the High Harmonic Generation Spectra," Phys. Rev. Lett. **80**, 3743–3746 (1998).

[18] K. Konishi, T. Higuchi, J. Li, J. Larsson, S. Ishii, and M. Kuwata-Gonokami, "Polarization-Controlled Circular Second-Harmonic Generation from Metal Hole Arrays with Threefold Rotational Symmetry," Phys. Rev. Lett. **112**, 135502, 135502 (2014).

[19] G. Li, S. Chen, N. Pholchai, B. Reineke, P. W. H. Wong, E. Y. B. Pun, K. W. Cheah, T. Zentgraf, and S. Zhang, "Continuous control of the nonlinearity phase for harmonic generations," Nat. Mater. **14**, 607–612 (2015).

[20] S. Chen, G. Li, F. Zeuner, W. H. Wong, E. Y. B. Pun, T. Zentgraf, K. W. Cheah, and S. Zhang, "Symmetry-Selective Third-Harmonic Generation from Plasmonic Metacrystals," Phys. Rev. Lett. **113**, 033901, 033901 (2014).

[21] S. Sioncke, T. Verbiest, and A. Persoons, "Second-order nonlinear optical properties of chiral materials," Mater. Sci. Eng. R **42**, 115–155 (2003).

[22] R. M. Hazen and D. S. Sholl, "Chiral selection on inorganic crystalline surfaces," Nat. Mater. **2**, 367–374 (2003).

[23] K.-H. Ernst, "Molecular chirality at surfaces," Phys. Status Solidi B **249**, 2057–2088 (2012).

[24] A. V. Andreev, A. A. Korneev, L. S. Mukina, M. M. Nazarov, I. R. Prudnikov, and A. P. Shkurinov, "Simultaneous generation of second and third optical harmonics on a metal grating," Phys. Rev. B **74**, 235421, 235421 (2006).

[25] J. I. Jang, S. Mani, J. B. Ketterson, P. Lovera, and G. Redmond, "Nonlinear refractive index and three-photon absorption coefficient of poly(9,9-dioctylfluorence)," Appl. Phys. Lett. **95**, 221906, 221906 (2009).

[26] N. Ueda, H. Kawazoe, Y. Watanabe, M. Takata, M. Yamane, and K. Kubodera, "Third-order nonlinear optical susceptibilities of electroconductive oxide thin films," Appl. Phys. Lett. **59**, 502–503 (1991).

[27] P. B. Johnson and R. W. Christy, "Optical constants of the noble metals," Phys. Rev. B **6**, 4370–4379 (1972).

[28] M. Campoy-Quiles, G. Heliotis, R. Xia, M. Ariu, M. Pintani, P. Etchegoin, and D. Bradley, "Ellipsometric characterization of the optical constants of polyfluorene gain media," Adv. Funct. Mater. **15**, 925–933 (2005).

[29] COMSOL, Inc., `http://www.comsol.com/`.

[30] K. O'Brien, H. Suchowski, J. Rho, A. Salandrino, B. Kante, X. Yin, and X. Zhang, "Predicting nonlinear properties of metamaterials from the linear response," Nat. Mater. **14**, 379–383 (2015).

[31] T. Heinz, "Nonlinear surface electromagnetic phenomena," in (Elsevier, Oxford, 1991) Chap. Second-Order Nonlinear Optical Effects at Surfaces and Interfaces.

Chapter 9

Enhanced magnetic second-harmonic generation from resonant metasurfaces

9.1 Introduction

The nonlinear optical properties of nanostructures are known to differ substantially from those of bulk media because they are affected by strong confinement and local resonances [1–7]. It is well established that the strong field enhancement through formation of 'hot spots' can dramatically boost nonlinear effects in metallic nanoparticles [8–11]. Importantly, in the case of metamaterials, the nano-patterning leads not only to more efficient nonlinear interaction, but also to completely new nonlinear regimes due to magnetic optical response of the constituent 'meta-atoms'. While exciting applications of the linear magnetic response of metamaterials for both metallic [12–16] and dielectric [17–19] structures have been readily achieved at infra-red and even optical frequencies, the magnetic nature of nonlinear optical phenomena in metamaterials is still largely unexplored.

We focus our attention on the process of second-harmonic generation, see Section 2.4. For small nanoparticles, efficient SHG may be observed due to several factors, including local field enhancement, deviation of the particle shapes from a symmetric one, surface effects, etc. [20]. Importantly, the resonances in plasmonic and dielectric nanoparticles, combined with a strong enhancement of the optical near-field, as well as the effective overlap of the interacting optical modes, allow for multi-fold enhancement of SHG [21–23].

It was recognized that the local field distribution depends drastically on the size, shape, and mutual orientation of the nanoparticles, thus providing a way for design of artificial materials with required nonlinear optical properties [24–26]. Such nonlinearity engineering resembles the engineering of nonlinear materials by chemical composition, where molecules of different symmetries can be selected to enable the required nonlinear polarization [27, 28]. However, the top-down approach enables the ultimate design of any arbitrary nonlinear polarization.

Most of the research in this field was performed for various types of (asymmetric)

plasmonic nanoparticles where the field localization effects are much more pronounced. Plasmonic metasurfaces exhibiting both electric and magnetic resonances have also been studied, aiming at obtaining magnetic nonlinear response from such metasurfaces. In particular, works on SHG in split-ring resonator metasurfaces [29, 30] have described the SHG signal as a contribution from the Lorentz force exerted on the electrons oscillating in the split-rings. However, the exact contribution of the magnetic nonlinearity was not quantified. Other works on fishnet metal-dielectric-metal metasurfaces have attributed the SHG signal predominantly to the surface SHG contribution, without the presence of magnetic nonlinearity [31]. These discrepancies have created a certain controversy about the nature of the observed harmonic generation from such resonant metasurfaces, namely if it arises from the electric or magnetic nonlinear polarization of the metasurfaces. Subsequent works have provided theoretical description of the physical mechanisms for achieving magnetic and electric nonlinear response [32, 33], however experimental verification in optics is still missing. In particular, the possibility to engineer the electric and magnetic nonlinear polarisabilities of the metasurface and the subsequent interference of second harmonic waves can have important implication for directional nonlinear generation as proposed earlier [34]. Therefore, further studies about the exact origin, electric or magnetic, of the process of SHG from resonant metasurfaces are required.

This chapter analyses experimentally and theoretically the SHG from an optical metasurface of metal-dielectric-metal nanodisks in the vicinity of optically-induced electric and magnetic resonances. We demonstrate that resonant excitation of different modes of nanoparticles allows for strong enhancement of SHG signal near the optically-induced magnetic resonance. The related theoretical analysis further shows that the SHG intensity is a result of the electric quadrupolar and magnetic dipolar contributions of the nonlinear material polarization.

The remainder of this chapter specifies the resonant nonlinear metasurface, which is at the heart of this study, and the setup for the experimental and computational investigation in Section 9.2. This is followed by the presentation of experimental and numerical results in Section 9.3. In addition, this section presents a theoretical formalism for the radiation by a periodic array of multipoles which is used to analyse the origin of the resonant enhancement of generated SH. The main conclusions of this study will finally be drawn in Section 9.4.

9.2 Experimental and simulation setup

This section specifies the design and geometry of a nonlinear metasurface with enhanced magnetic second-harmonic generation, explains fabrication, measurement and how it can be modelled numerically.

9.2.1 Device specification and experiment

For our experiments, we employ metasurfaces consisting of a regular array of metal-dielectric-metal disk-like nanoparticles (meta-atoms). We use two metal disks separated by a dielectric spacer to obtain two distinct resonant modes of the meta-atom excited by incident light at two different wavelengths [35, 36]. The two resonances are associated with co- and counter-propagating electric currents excited in the parallel metal disks. In particular, the counter-propagating electric currents result in non-zero electric quadrupole and magnetic dipole linear response in the meta-atoms are therefore referred below as magnetic resonances of the metasurface.

The SEM image of the fabricated metasurface is shown in Fig. 9.1(a). The inset shows a side view of a single meta-atom. The samples are fabricated on a glass substrate using an electron-beam lithography. The glass substrate is covered by a $10\,\mathrm{nm}$ thin conductive ITO layer to prevent charging during the lithography. This is followed by an e-beam evaporation of Au, MgF_2 and Au layers and by a sequential lift-off procedure. The size of the patterned area is $100\,\mu\mathrm{m} \times 100\,\mu\mathrm{m}$. In our sample, the thicknesses of the two Au disks and MgF_2 layer are $30\,\mathrm{nm}$. The meta-atom has a bottom diameter of $140\,\mathrm{nm}$, with a small tapering angle of 10° due to the evaporation procedure. The meta-atoms are arranged into a square lattice with a $400\,\mathrm{nm} \times 400\,\mathrm{nm}$ unit cell. The sample has been fabricated and characterised by SEM imaging by S. Kruk at the Australian National University.

We study the second-order optical response of the metasurface for both TM and TE polarizations of the pump beam and different angles of incidence φ, as shown schematically in Figs. 9.1(b) and 9.1(c). The sample is illuminated from the top side (fundamental beam spot size of $50\,\mu\mathrm{m}$) and the TM polarized second harmonic signal is collected from the substrate side. The SHG spectroscopy is studied experimentally by employing a Titanium-Sapphire laser ($100\,\mathrm{fs}$ pulses, $80\,\mathrm{MHz}$ repetition rate, tuning range from $720\,\mathrm{nm}$ to $855\,\mathrm{nm}$,

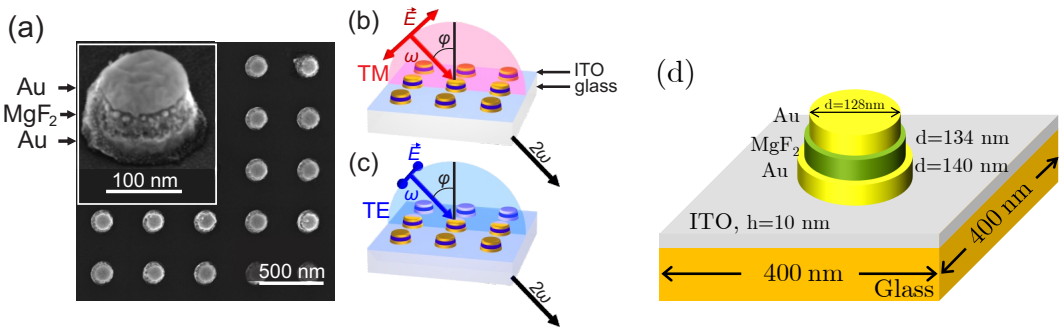

Figure 9.1: a) SEM image of a fabricated metasurface. Inset shows a side view of a single three-layer meta-atom made of $Au/MgF_2/Au$ layers [The sample has been fabricated and characterised by SEM imaging by S. Kruk at the Australian National University.]. b) and c) Geometry of the SHG experiment for TM- and TE-polarized pump waves, respectively. d) Geometry of the unit cell used in the simulations of the device.

peak intensity in the focal plane $0.3\,\mathrm{GW\,cm^{-2}}$) as a source of the fundamental radiation. The SHG radiation transmitted through the sample is collected by a lens with 0.17 NA and is spectrally selected by a set of colour filters and a monochromator. The SHG signal is detected by spectral measurements and exhibits a square-law dependence on the pump intensity. In addition, we verified that the spectra of the nonlinear signal correspond to the spectra of second harmonic. The nonlinear measurements have been performed by A. Bykov et al. at the Lomonosov Moscow State University.

9.2.2 Simulation method

The linear optical behaviour of the metasurface is calculated by means of the RCWA. The tapering angle of 10° is not directly taken into account. Instead, three RCWA layers containing straight-walled Au-, MgF_2- and Au-cylinders with diameters of $128\,\mathrm{nm}$, $134\,\mathrm{nm}$, and $140\,\mathrm{nm}$, respectively, are used in the simulations, as per Fig. 9.1(d). Additional and independently, the linear transmission spectra have also been calculated by S. Kruk at the Australian National University by using CST Microwave Studio [37] and they agree well with the RCWA calculations.

The nonlinear optical response of the array of meta-atoms at the SH comprises several possible contributions, namely the surface (local) and bulk (nonlocal) SHG from the metallic nanodisks, as well as bulk (local) SHG from metallic disks and $\mathrm{MgF_2}$ dielectric layer. However, because of the centrosymmetric nature of the Au crystalline lattice and polycrystalline structure of $\mathrm{MgF_2}$, the last two contributions can be neglected. As a result, the sources of the SH field can be described by the following nonlinear polarizations [38],

$$\mathbf{P}^s_\Omega(\mathbf{r}) = \epsilon_0 \hat{\chi}^{(2)}_s : \mathbf{E}_\omega(\mathbf{r})\mathbf{E}_\omega(\mathbf{r})\delta(\mathbf{r}-\mathbf{r}_s), \tag{9.1a}$$

$$\mathbf{P}^b_\Omega(\mathbf{r}) = \epsilon_0\delta'(\mathbf{E}_\omega(\mathbf{r})\cdot\nabla)\mathbf{E}_\omega(\mathbf{r}) + \epsilon_0\beta\mathbf{E}_\omega(\mathbf{r})(\nabla\cdot\mathbf{E}_\omega(\mathbf{r})) + \epsilon_0\gamma\nabla(\mathbf{E}_\omega(\mathbf{r})\cdot\mathbf{E}_\omega(\mathbf{r})), \tag{9.1b}$$

where $\mathbf{P}^s_\Omega$ and $\mathbf{P}^b_\Omega$ describe the surface and bulk nonlinear polarization, respectively, and $\mathbf{r}_s$ defines the surface. The surface of metallic parts of the meta-atom possesses an isotropic mirror-symmetry plane perpendicular to the surface, so that the surface second-order susceptibility tensor, $\hat{\chi}^{(2)}_s$, has only three independent components, $\hat{\chi}^{(2)}_{s,\perp\perp\perp}$, $\hat{\chi}^{(2)}_{s,\perp\parallel\parallel}$, and $\hat{\chi}^{(2)}_{s,\parallel\parallel\perp} = \hat{\chi}^{(2)}_{s,\parallel\perp\parallel}$, where the symbols $\perp$ and $\parallel$ refer to the directions normal and tangent to the surface, respectively. For gold, the corresponding values are [39]: $\hat{\chi}^{(2)}_{s,\perp\perp\perp}$ = -$(0.86+1.34\mathrm{i})\times10^{-18}\mathrm{m^2\,V^{-1}}$, $\hat{\chi}^{(2)}_{s,\parallel\parallel\perp}$ = -$(4.61+0.43\mathrm{i})\times10^{-20}\mathrm{m^2\,V^{-1}}$. $\hat{\chi}^{(2)}_{s,\perp\parallel\parallel}$ can be neglected as compared with other components, whereas the bulk susceptibility is $\gamma = 7.13\times10^{-21}\,\mathrm{m^2\,V^{-1}}$. Moreover, the dominant contribution to the bulk polarization comes from the third term in Eq. (9.1b), as in homogeneous media the second term vanishes and the ratio between the first and third terms is about ν/ω, where ν is the damp-

ing frequency (at optical frequencies, $\nu \ll \omega$) [40]; therefore, in our simulations we set $\delta' = \beta = 0$. Note that we did not include in Eq. (9.1b) a fourth term that accounts for the anisotropy of the medium because in our system the nonlinear bulk polarization is due entirely to the free electrons, namely it is generated in an isotropic medium.

To calculate the response at the SH frequency, the nonlinear GSM is used. In contrast to its formulation in Chapter 7 it is modified in two ways: The first of the three computational steps in Section 7.2.1, namely the linear calculation, is performed by the RCWA to obtain the electric field, $\mathbf{E}_\omega$, at the FF frequency. In the second step, nonlinear SH polarization Eq. (9.1) is calculated from $\mathbf{E}_\omega$. In the third step, this nonlinear polarization is added to the linear polarization at the SH and the linear GSM is used to compute the electromagnetic fields at the SH frequency as well as the corresponding transmission and reflection coefficients. Simulations have been performed with $N = 18$ harmonics and each GSM layer had a width of $1\,\mathrm{nm}$.

9.3 Experimental and theoretical insights into second-harmonic generation from a resonant metasurface

9.3.1 Linear and nonlinear resonances in gold nanodisks

First, we measure the linear transmission of the metasurface at normal incidence, as shown in Figure 9.2(a), grey curve. These linear measurements have been performed by S. Kruk at the Australian National University. The experimental spectra feature two resonances centred at $640\,\mathrm{nm}$ and $790\,\mathrm{nm}$ wavelengths. The spectral position of these resonances in numerically calculated transmission spectrum in Fig. 9.3(a) agrees well with the experimental results. The numerically calculated resonances around $640\,\mathrm{nm}$ exhibits a stronger decrease of transmission, whereas the RCWA underestimates the strengths of the second resonance around $790\,\mathrm{nm}$. The differences can be explained in two ways: First, the actual sample differs from its specification due to potential shortcomings of the nano-fabrication techniques. And secondly, the tapering angle was not directly taken into account in the simulations, as was explained in Section 9.2.2.

Figure 9.2(a) shows the spectra of the SH intensity measured at the incidence angle $\varphi = 30^\circ$ for two linear polarizations of the fundamental laser beam. Both SH spectra show a pronounced efficiency enhancement near the resonance at $790\,\mathrm{nm}$ wavelength. Figures 9.2(b) and (c) show the angular dependence of the SH spectra for the two polarizations. We note that at normal incidence the SHG efficiency nearly vanishes, as well as the SHG from the substrate remains below the noise level for all angles of incidence. The vanishing SHG at normal incidence is expected because the design of the metasurface elements is close to centrosymmetric, even with the presence of a substrate and a tilting shape of the meta-atoms, and hence the SHG process is almost prohibited. In contrast, for oblique

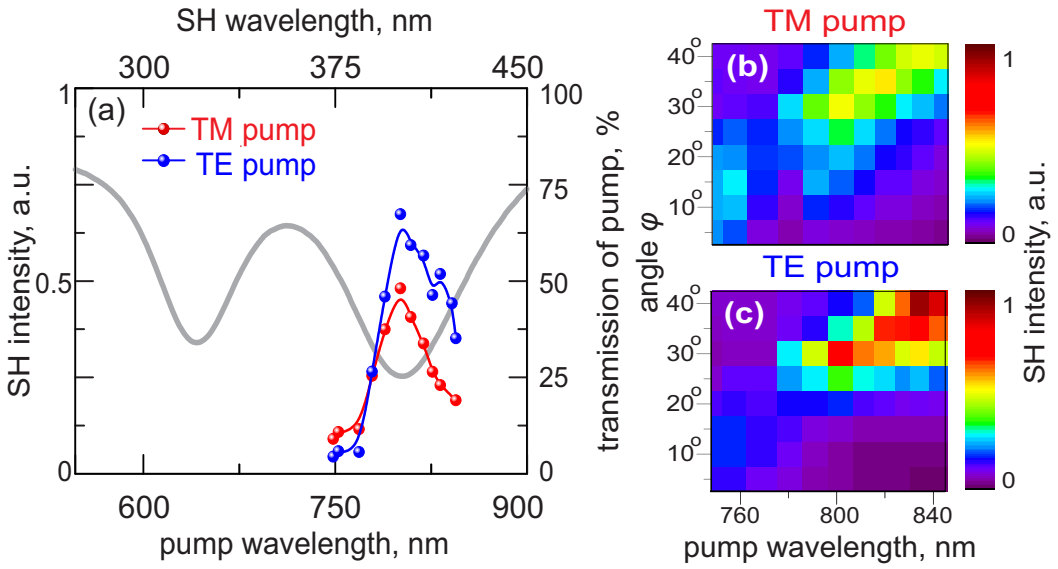

Figure 9.2: Experimental results. a) Linear transmission of the metasurface (grey) at normal incidence and second harmonic spectra for TM- and TE-polarized pump waves (red and blue) measured for the incident angle of 30°. [The linear measurements have been performed by S. Kruk at the Australian National University.] b and c) Second-harmonic intensity vs. pump wavelength and angle of incidence for TM and TE polarizations, respectively. [The nonlinear measurements have been performed by A. Bykov et al. at the Lomonosov Moscow State University.]

incidence surface polarisations contribute effectively to the frequency conversion, and the SHG process becomes more efficient with an increase of the incident angle φ. In addition, we observe a red-shift of the SH spectral maxima with an increase of the incidence angle φ for both TE and TM polarizations of the fundamental beam.

In order to gain a deeper insight into our experimental results and we first calculate the linear transmission for normal and 30° oblique incidence, see Fig. 9.3(a). The positions of the resonances in the calculated spectra are in a good agreement with the experimental data, corresponding to an electric resonance at approximately $640\,\mathrm{nm}$ and a magnetic one at $790\,\mathrm{nm}$. The transmission spectra at normal incidence are qualitatively similar to the spectra at oblique incidence. A small red-shift of the resonant wavelength can be observed for the oblique excitation. At the first resonance near $640\,\mathrm{nm}$, co-propagating currents are excited in the top and bottom metallic nanodisks of the meta-atom. At the second resonance, counter-propagating currents are excited in the two metallic nanodisks, as shown in Fig. 9.3(c).

The SH spectra for two linear polarisations of pump determined by this approach, are shown in Fig. 9.3(b). Both SHG theoretical spectra are in a good agreement with the experimental data, and they show pronounced maxima near the second (magnetic) resonance at around $790\,\mathrm{nm}$. Importantly, the theory predicts less efficient SHG near the first (electric) resonance at $640\,\mathrm{nm}$.

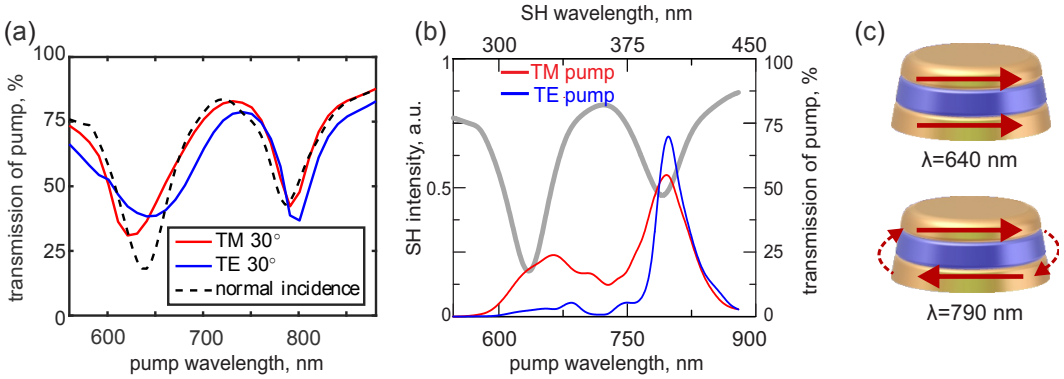

Figure 9.3: Numerical results: a) Numerical transmission spectrum of the metasurface at normal incidence (dashed black) and SH spectra for TM- and TE-polarized pumps (red and blue, respectively) calculated for the incidence angle $\phi = 30°$. b) SH radiation spectrum in direction of transmission for TM- and TE-polarized pumps (red and blue, respectively). The grey line indicates the linear transmission at the FF. c) Schematic of the electric currents distribution in a single meta-atom at the first and second resonant wavelengths. Dashed lines visualize the displacement current.

9.3.2 Theory of multipolar expansion of the electromagnetic fields

Although the linear and nonlinear experiments and simulations already give some insight into the nature of these two resonances, it is best revealed by applying the multipolar expansion for the sources of the scattered field. The electric charges and currents of the meta-atom give rise to electromagnetic multipoles. The first three terms in the multipolar expansion, the electric dipole, magnetic dipole, and electric quadrupole, have the definitions [41, 42],

$$\mathbf{p}_\omega = \int \mathbf{P}_\omega(\mathbf{r}) d\mathbf{r}, \tag{9.2a}$$

$$\mathbf{m}_\omega = -i\frac{\omega}{2}\int \mathbf{r} \times \mathbf{P}_\omega(\mathbf{r}) d\mathbf{r}, \tag{9.2b}$$

$$Q_{\omega,jk} = \int [3(r_j P_{\omega,k} + r_k P_{\omega,j}) - 2\delta_{jk}\mathbf{r}\cdot\mathbf{P}_\omega(\mathbf{r})] d\mathbf{r}. \tag{9.2c}$$

Here $\mathbf{P}_\omega(\mathbf{r}) = \epsilon_0[\epsilon_r(\omega) - 1]\mathbf{E}_\omega(\mathbf{r})$ is the polarization at the location point $\mathbf{r}$, with $\mathbf{E}_\omega(\mathbf{r})$ being the electric field and ϵ_r the relative electric permittivity, and $\mathbf{p}_\omega$, $\mathbf{m}_\omega$, and $Q_{\omega,jk}$ are the electric dipole, magnetic dipole, and electric quadrupole of the nanodisk meta-atom, respectively. Importantly, Eq. (9.2) include the contributions of both displacement and conductive currents if the frequency dispersion of ϵ_r is taken into account. We notice that, since the contribution of the substrate into the far-field is shown to be negligible within the experimental accuracy, the integration in Eq. (9.2) is restricted to the volume of the meta-atom.

9.3.3 Radiation by a periodic array of multipoles

The magnitude of multipoles can be misleading when quantifying their respective influence on the total scattered field due to the interference among the radiation emitted by different

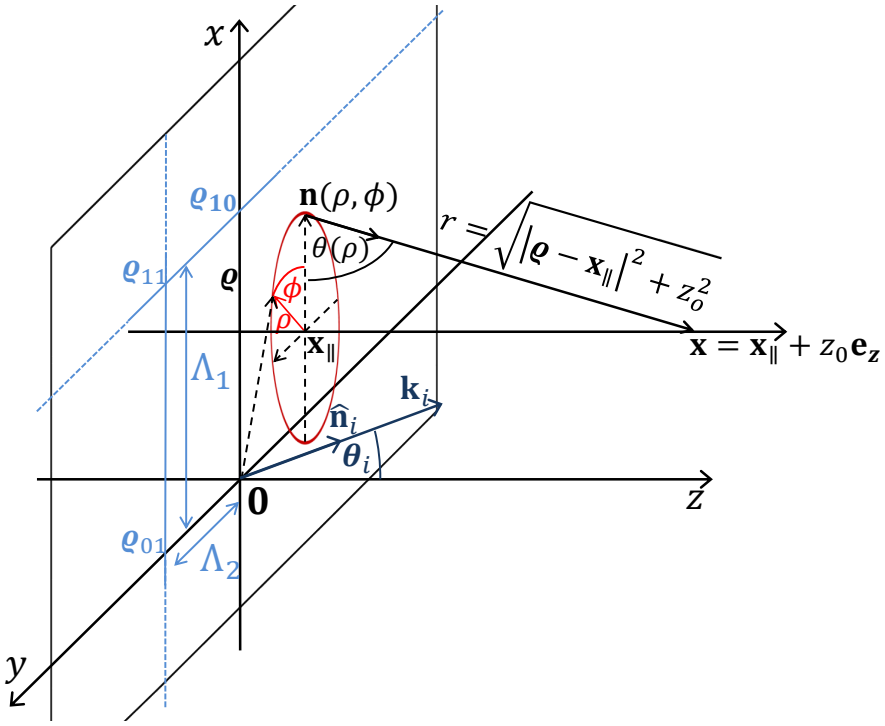

Figure 9.4: Schematics of the wave scattering configuration. The lattice of multipoles and the Cartesian system in which the scattered fields are calculated have a common origin. The observation point is located at a distance z_0 from the plane of multipoles, the distance between its projection onto this plane and the origin of the lattice being $\mathbf{x}_\parallel$.

multipoles and because not individual but a periodic array of multipoles is considered.

Therefore, in this section we will derive the formula for the intensity of light emitted by a two-dimensional periodic distribution of multipoles, namely electric dipoles, magnetic dipoles, and electric quadrupoles. We will present in detail only the case of a lattice of electric dipoles, in the remaining two cases only giving the final formula for the wave intensity. Thus, the periodic lattice is defined by the primitive vectors $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$, meaning that they are located at $\boldsymbol{\varrho}_{n_1n_2} = n_1\mathbf{\Lambda}_1 + n_2\mathbf{\Lambda}_2$, where n_1 and n_2 are arbitrary integers. We assume that a plane wave with wavenumber $k = \frac{\omega}{c} = \frac{2\pi}{\lambda}$, where ω and λ are the frequency and wavelength of the wave, respectively, propagates along a direction defined by the unit vector, $\hat{\mathbf{n}}_i$, which makes an angle of incidence, θ_i, with the normal to the plane of multipoles. Since the electric dipoles are generated by the incoming plane wave, one can write $\mathbf{p}(\boldsymbol{\varrho}_{n_1n_2}) = \mathbf{p}(0)\exp(ik\hat{\mathbf{n}}_{i,\parallel}\cdot\boldsymbol{\varrho}_{n_1n_2})$, where $\mathbf{p}(\mathbf{x})$ is the dipole at location $\mathbf{x}$ and $\hat{\mathbf{n}}_{i,\parallel}$ is the projection of $\hat{\mathbf{n}}_i$ onto the plane of multipoles.

The electric far-field at position $\mathbf{x}$, generated by an electric dipole located at $\mathbf{x}'$ (see Fig. 9.4), is

$$\mathbf{E}_p(\mathbf{x},\mathbf{x}') = \frac{\mu_0\omega^2}{4\pi}[\hat{\mathbf{n}}\times\mathbf{p}(\mathbf{x}')]\times\hat{\mathbf{n}}\frac{e^{ikr}}{r}, \tag{9.3}$$

where $\mathbf{r} = \mathbf{x} - \mathbf{x}' = r\hat{\mathbf{n}}$. If the wavelength of the incident plane wave is much larger than

the unit cell of the lattice the total electric field can be calculated by replacing the sum over the electric dipoles with an integral over the plane of the lattice,

$$\mathbf{E}_p(\mathbf{x}) = \sum_{n_1,n_2} \mathbf{E}_p(\mathbf{x}, \boldsymbol{\varrho}_{n_1 n_2}) \approx \frac{1}{A_0} \int_{\mathbb{R}^2} \mathbf{E}_p(\mathbf{x}, \boldsymbol{\varrho}) d\boldsymbol{\varrho}, \tag{9.4}$$

where $A_0 = \|\boldsymbol{\Lambda}_1 \times \boldsymbol{\Lambda}_2\|$ is the area of the unit cell of the lattice. Without loss of generality, we can choose $\phi_i = 0$, i.e. $\mathbf{k}_i = k(\sin\theta_i, 0, \cos\theta_i)$, meaning that $\hat{\mathbf{n}}_i = (\sin\theta_i, 0, \cos\theta_i)$. Then, by changing the integration variables to a cylindrical coordinate system centred at the position of the projection of the observation point onto the plane of the lattice of dipoles, the far-field in Eq. (9.4) can be expressed as:

$$\mathbf{E}_p(\mathbf{x}_\parallel + z_0\hat{\mathbf{e}}_z) = \frac{\mu_0\omega^2}{4\pi A_0} \int_{\mathbb{R}^2} (\hat{\mathbf{n}} \times \mathbf{p}_0) \times \hat{\mathbf{n}} \frac{1}{\sqrt{\rho^2 + z_0^2}} e^{ik\left[\sqrt{\rho^2+z_0^2} + \hat{\mathbf{n}}_{i,\parallel}\cdot(\boldsymbol{\rho}+\mathbf{x}_\parallel)\right]} d\boldsymbol{\rho}, \tag{9.5}$$

where $\boldsymbol{\varrho} = \mathbf{x}_\parallel + \boldsymbol{\rho}$, $\mathbf{x}_\parallel$ being the distance between the origin of the lattice and the projection of the observation point onto the lattice plane, and $\mathbf{p}_0 = \mathbf{p}(0)$. Expressing $\hat{\mathbf{n}} = (\cos\theta\cos\phi, \cos\theta\sin\phi, \sin\theta)$ and $\boldsymbol{\rho} = (\rho\cos\phi, \rho\sin\phi, 0)$, this integral is transformed to

$$\mathbf{E}_p(\mathbf{x}) = \frac{\mu_0\omega^2}{4\pi A_0} e^{ik\hat{\mathbf{n}}_{i,\parallel}\cdot\mathbf{x}_\parallel} \int_0^{2\pi}\int_0^{\infty} (\hat{\mathbf{n}} \times \mathbf{p}_0) \times \hat{\mathbf{n}} \frac{\rho}{\sqrt{\rho^2 + z_0^2}} e^{ik\left(\sqrt{\rho^2+z_0^2} + \rho\sin\theta_i\cos\phi\right)} d\rho d\phi. \tag{9.6}$$

We now make another change of variable, $\rho(\theta) = z_0\cot\theta$, so that (9.6) becomes

$$\begin{aligned}\mathbf{E}_p(\mathbf{x}) &= \frac{\mu_0\omega^2}{4\pi A_0} e^{ik\hat{\mathbf{n}}_{i,\parallel}\cdot\mathbf{x}_\parallel} \int_0^{2\pi}\int_0^{\pi} (\hat{\mathbf{n}} \times \mathbf{p}_0) \times \hat{\mathbf{n}} \frac{z_0\cos\theta}{\sin^2\theta} e^{ikz_0(\csc\theta + \sin\theta_i\cot\theta\cos\phi)} d\theta d\phi \\ &\equiv \frac{\mu_0\omega^2}{4\pi A_0} e^{ik\hat{\mathbf{n}}_{i,\parallel}\cdot\mathbf{x}_\parallel} \int_0^{2\pi}\int_0^{\pi} g(\theta,\phi;z_0) e^{ikz_0 f(\theta,\phi)} d\theta d\phi,\end{aligned} \tag{9.7}$$

where $g(\theta,\phi;z_0) = z_0\frac{\cos\theta}{\sin^2\theta}[\hat{\mathbf{n}}(\theta,\phi)\times\mathbf{p}_0]\times\hat{\mathbf{n}}(\theta,\phi)$ and $f(\theta,\phi) = \csc\theta + \sin\theta_i\cot\theta\cos\phi$.

Since in the far-field region $z_0 \to \infty$, the integral Eq. (9.7) can be evaluated using the multi-dimensional form of the stationary phase approximation formula (see, e.g., Ref. [43]), the result being

$$\mathbf{E}_p(\mathbf{x}) \approx \frac{\mu_0\omega^2}{4\pi A_0} e^{ik\hat{\mathbf{n}}_{i,\parallel}\cdot\mathbf{x}_\parallel} \frac{2\pi}{kz_0} g(\tilde{\theta},\tilde{\phi};z_0) e^{ikz_0 f(\tilde{\theta},\tilde{\phi})} \left[f_{\theta\theta}(\tilde{\theta},\tilde{\phi}) f_{\phi\phi}(\tilde{\theta},\tilde{\phi})\right]^{-\frac{1}{2}} e^{\frac{i\pi}{4}\sigma(\tilde{\theta},\tilde{\phi})}, \tag{9.8}$$

where $(\tilde{\theta},\tilde{\phi}) = (\pi/2 - \theta_i, \pi)$ is the only stationary point of f with a nonvanishing contribution to the integral and the constant $\sigma(\tilde{\theta},\tilde{\phi}) = \operatorname{sgn}\left(f_{\theta\theta}(\tilde{\theta},\tilde{\phi})\right) + \operatorname{sgn}\left(f_{\phi\phi}(\tilde{\theta},\tilde{\phi})\right) = 2$ for $0 < \theta_i < \frac{\pi}{2}$. After simple calculations, the integral Eq. (9.7) can hence be cast into the

final form:

$$\mathbf{E}_p(\mathbf{x}) = \frac{i\omega Z_0}{2A_0\cos\theta_i}(\hat{\mathbf{n}}_i \times \mathbf{p}_0) \times \hat{\mathbf{n}}_i e^{ik(x\sin\theta_i + z_0\cos\theta_i)} \equiv \mathbf{E}_{p0} e^{ik\hat{\mathbf{n}}_i\cdot\mathbf{x}}, \tag{9.9}$$

i.e. the far-field is a plane wave with amplitude, $\mathbf{E}_{p0} = \frac{i\omega Z_0}{2A_0\cos\theta_i}(\hat{\mathbf{n}}_i \times \mathbf{p}_0) \times \hat{\mathbf{n}}_i$, and direction of propagation, $\hat{\mathbf{n}}_i$. Consequently, the corresponding Poynting vector is:

$$\mathbf{S}_p(\mathbf{x}) = \frac{1}{2}\mathsf{Re}\left\{\mathbf{E}_p(\mathbf{x}) \times \mathbf{H}_p^*(\mathbf{x})\right\} = \frac{Z_0\omega^2}{8A_0^2\cos^2\theta_i}|(\hat{\mathbf{n}}_i \times \mathbf{p}_0) \times \hat{\mathbf{n}}_i|^2\hat{\mathbf{n}}_i. \tag{9.10}$$

In a similar way one can calculate the wave intensity emitted by lattices of magnetic dipoles and electric quadrupoles. Thus, starting from the corresponding far-field expressions,

$$\mathbf{E}_m(\mathbf{x},\mathbf{x}') = -\frac{1}{c}\frac{\mu_0\omega^2}{4\pi}\hat{\mathbf{n}} \times \mathbf{m}(\mathbf{x}')\frac{e^{ikr}}{r}, \tag{9.11a}$$

$$\mathbf{E}_Q(\mathbf{x},\mathbf{x}') = -\frac{ik}{6}\frac{\mu_0\omega^2}{4\pi}[\hat{\mathbf{n}} \times \mathbf{Q}(\mathbf{x}';\hat{\mathbf{n}})] \times \hat{\mathbf{n}}\frac{e^{ikr}}{r}, \tag{9.11b}$$

one can easily derive the formulae for the corresponding radiated far-fields:

$$\mathbf{E}_m(\mathbf{x}) = -\frac{1}{c}\frac{i\omega Z_0}{2A_0\cos\theta_i}\hat{\mathbf{n}}_i \times \mathbf{m}_0 e^{ik\hat{\mathbf{n}}_i\cdot\mathbf{x}} \equiv \mathbf{E}_{m0}e^{ik\hat{\mathbf{n}}_i\cdot\mathbf{x}}, \tag{9.12a}$$

$$\mathbf{E}_Q(\mathbf{x}) = -\frac{ik}{6}\frac{i\omega Z_0}{2A_0\cos\theta_i}[\hat{\mathbf{n}}_i \times \mathbf{Q}_0(\hat{\mathbf{n}}_i)] \times \hat{\mathbf{n}}_i e^{ik\hat{\mathbf{n}}_i\cdot\mathbf{x}} \equiv \mathbf{E}_{Q0}e^{ik\hat{\mathbf{n}}_i\cdot\mathbf{x}}. \tag{9.12b}$$

Combining results Eq. (9.10) and Eq. (9.12) one obtains the formula for the wave intensity radiated at the fundamental frequency by the lattice of electric dipoles, magnetic dipoles, and electric quadrupoles:

$$I_\omega = \frac{Z_0\omega^2}{8A_0^2\cos^2\theta_i}\left|(\hat{\mathbf{n}}_i \times \mathbf{p}_0) \times \hat{\mathbf{n}}_i - \frac{1}{c}\hat{\mathbf{n}}_i \times \mathbf{m}_0 - \frac{ik}{6}[\hat{\mathbf{n}}_i \times \mathbf{Q}_0(\hat{\mathbf{n}}_i)] \times \hat{\mathbf{n}}_i\right|^2. \tag{9.13}$$

Note that if one makes the changes $\omega \to \Omega = 2\omega$ and $k \to K = \Omega/c = 2k$, Eq. (9.13) can be used to calculate the field intensity radiated at the second harmonic as well. This is an immediate consequence of the quadratic dependence on the fundamental field of the polarization at the second harmonic. Thus, at the second harmonic the electric dipoles depend on their location in the lattice via $\mathbf{p}_\Omega(\boldsymbol{\varrho}_{n_1n_2}) = \mathbf{p}_\Omega(0)e^{2ik\hat{\mathbf{n}}_{i,\parallel}\cdot\boldsymbol{\varrho}_{n_1n_2}} = \mathbf{p}_{\boldsymbol{\Omega}}(0)e^{iK\hat{\mathbf{n}}_{i,\parallel}\cdot\boldsymbol{\varrho}_{n_1n_2}}$, and the same relation holds for the magnetic dipoles and electric quadrupoles. Therefore, the field at the second harmonic is radiated along the direction of the incident wave, the corresponding intensity, I_Ω, being given by Eq. (9.13) with $\omega \to \Omega$ and $k \to K$.

9.3.4 Multipolar analysis of the resonant metasurface

We perform the multipolar expansion using Eq. (9.2) that provides us with a quantitative analysis of the sources of the far-field at the pump by comparing their radiated intensity according to Eq. (9.13). The results of this analysis are summarized in Figs. 9.5(a) and 9.5(b), and they suggest that the first resonance originates from the excitation of an electric dipole, whereas the second spectral peak corresponds to the excitation of a mixture of magnetic dipole and electric quadrupole moments resonantly excited together with the electric dipole. In addition, it can be seen that, for both types of the pump polarization, for wavelengths smaller than the wavelength of the magnetic resonance, the radiation emitted by the multipoles interfere constructively so that the total radiated power is higher than the power produced solely by the electric dipole. In the spectral region of the magnetic resonance and beyond, destructive interference is observed and the total radiated power becomes lower than that of the electric dipole alone.

To understand the difference in the SHG efficiency observed at the two resonances, we employ a multipole expansion of the sources of the nonlinear field using a similar approach to that used at the FF. Specifically, we calculate numerically the distributions of nonlinear electric currents and charges induced in the meta-atoms by the optical pump and, subsequently, by using Eqs. (9.2) and (9.1), the nonlinear multipoles $\mathbf{p}_\Omega$, $\mathbf{m}_\Omega$, and $Q_{\Omega,ij}$ associated with these distributions [42, 44]. We emphasize that in these calculations we consider only the nonlinear source polarizations given in Eq. (9.1), with the linear polarization of the medium at the SH frequency being excluded. Consequently, the resonant effects due to geometrical and material characteristics of the meta-atom on the radiative properties of these multipoles are not incorporated into our analysis, which implies that the multipoles are assumed to radiate as if they were in vacuum. By doing this, we want to separate the purely nonlinear properties of those multipoles from the linear properties in the spectral region of the second-harmonic generation. We follow this procedure to calculate the spectra of the first three terms of the multipolar expansion Eq. (9.2) induced in the meta-atom by a pump wave incident onto the metasurface at the angle $\varphi = 30°$.

Starting from the nonlinear optical response of a single meta-atom, we can now proceed to study the nonlinear properties of a metasurface created by a periodic square lattice of such meta-atoms as described by Eq. (9.13). The power spectra corresponding to the three types of multipoles, as well as the spectrum of the total emitted power are presented in Figs. 9.5(c) and 9.5(d) for both TM and TE polarized pump, respectively. While there are quantitative differences between these SHG spectra and those obtained using rigorous calculations based on the GSM, see Fig. 9.3(b), both methods predict the same qualitative features of the SHG spectra, namely two spectral peaks corresponding to the electric and magnetic resonances, which are separated by a spectral region of weak SHG. More

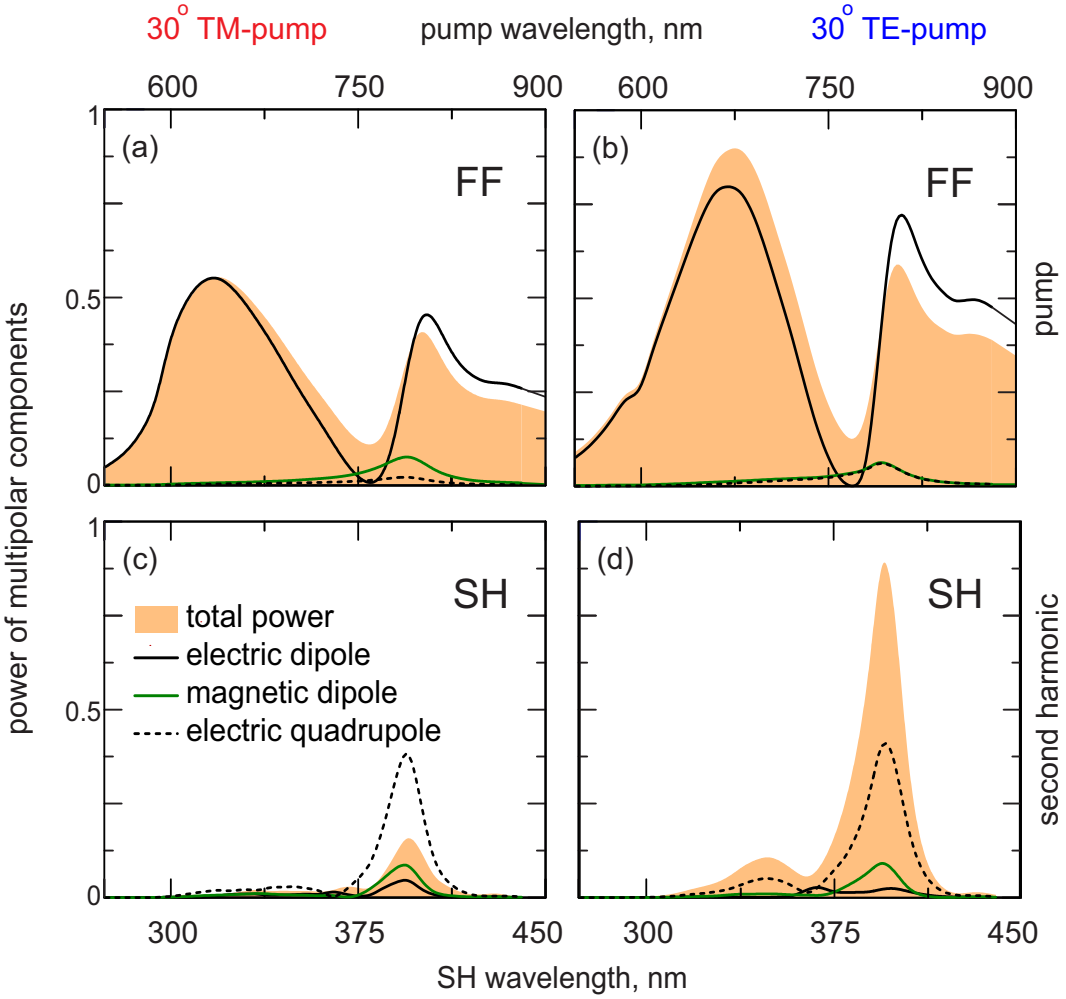

Figure 9.5: Numerically calculated radiation of the electric dipole (black, solid), electric quadrupole (black, dashed), and magnetic dipole components (green) at (a,b) fundamental wavelengths and (c,d) second-harmonic wavelengths, for two polarisations of the pump at the oblique incidence.

specifically, the radiated SHG intensity is dominated by the higher-order multipoles, electric quadrupole and magnetic dipole, with less than 3% contribution from the electric dipole.

Note that the spectral location of this weak SHG region, which can also be observed in the experimentally measured spectra, as per Figs. 9.2(b) and 9.2(c), is essentially independent of the angle of incidence of the FF beam. This observation confirms our theoretical predictions of a vanishing nonlinear dipole moment at this particular frequency. In addition, similarly to the linear optical response of the metasurface, these spectra illustrate that the nonlinear properties of the metasurface are determined by subtle interference effects among the radiation emitted by the three different types of multipoles, as observed in Figs. 9.5(c) and 9.5(d).

9.4 Conclusions

In this chapter, the second-harmonic generation from a resonant nonlinear metasurface composed of metal-dielectric-metal nanodisks has been studied experimentally and numerically, and the measured values and the results calculated by the RCWA and the nonlinear GSM show good agreement. In addition, a multipolar analysis has been performed to identify the electric and magnetic nature of its resonances. Importantly, the nonlinear radiation at the higher wavelength, magnetic resonance is chiefly due to the excitation of higher-order multipoles, namely the magnetic dipole and electric quadrupole, which are induced by the pump wave at oblique incidence. This theoretical analysis is based on the numerical results obtained by the RCWA and the nonlinear GSM, which emphasises their relevance and potential for the investigation and explanation of complex nonlinear phenomena in optical diffraction gratings and metasurfaces. We believe these results will pave a way to establishing novel engineered nonlinear materials with enhanced harmonic efficiency, driven by optically-induced magnetic resonances. This will find immediate applications in colloidal chemistry, surface science, and catalytic chemistry [38, 45–47].

Bibliography

[1] U. Kreibig and M. Vollmer, *Optical properties of metal clusters* (Springer-Verlag, Berlin, 1995).

[2] W. L. Barnes, A. Dereux, and T. W. Ebbesen, "Surface plasmon subwavelength optics," Nature **424**, 824 (2003).

[3] A. V. Zayats, I. I. Smolyaninov, and A. A. Maradudin, "Nano-optics of surface plasmon polaritons," Phys. Reports **408**, 131 (2005).

[4] S. A. Maier, *Plasmonics: Fundamentals and Applications* (Springer, New York, 2007).

[5] N. C. Panoiu and R. M. Osgood, "Subwavelength nonlinear plasmonic nanowire," Nano Lett. **4**, 2427–2430 (2004).

[6] J. Y. Suh and T. W. Odom, "Nonlinear properties of nanoscale antennas," Nano Today **8**, 469 –479 (2013).

[7] M. R. Shcherbakov, D. N. Neshev, B. Hopkins, A. S. Shorokhov, I. Staude, E. V. Melik-Gaykazyan, M. Decker, A. A. Ezhov, A. E. Miroshnichenko, I. Brener, A. A. Fedyanin, and Y. S. Kivshar, "Enhanced Third-Harmonic Generation in Silicon Nanoparticles Driven by Magnetic Response," Nano Lett. **14**, 6488–6492 (2014).

[8] J. A. H. Van Nieuwstadt, M Sandtke, R. H. Harmsen, F. B. Segerink, J. C. Prangsma, S. Enoch, and L Kuipers, "Strong modification of the nonlinear optical response of metallic subwavelength hole arrays," Phys. Rev. Lett. **97**, 146102 (2006).

[9] W. Fan, S. Zhang, N.-C. Panoiu, A Abdenour, S Krishna, R. Osgood, K. Malloy, and S. Brueck, "Second harmonic generation from a nanopatterned isotropic nonlinear material," Nano Lett. **6**, 1027–1030 (2006).

[10] M. Kauranen and A. V. Zayats, "Nonlinear plasmonics," Nat. Photon. **6**, 737–748 (2012).

[11] M. Celebrano, X. Wu, M. Baselli, S. Gromann, P. Biagioni, A. Locatelli, C. De Angelis, G. Cerullo, R. Osellame, B. Hecht, L. Du, F. Ciccacci, and M. Finazzi, "Mode matching in multiresonant plasmonic nanoantennas for enhanced second harmonic generation," Nat Nano **10**, 412–417 (2015).

[12] J. B. Pendry, A. J. Holden, D. J. Robbins, and W. J. Stewart, "Magnetism from conductors and enhanced nonlinear phenomena," IEEE Trans. Microwave Theory Tech. **47**, 2075 (1999).

[13] S. Linden, C. Enkrich, M. Wegener, J. F. Zhou, T. Koschny, and C. M. Soukoulis, "Magnetic response of metamaterials at 100 terahertz," Science **306**, 1351–1353 (2004).

[14] T. J. Yen, W. J. Padilla, N. Fang, D. C. Vier, D. R. Smith, J. B. Pendry, D. N. Basov, and X. Zhang, "Terahertz magnetic response from artificial materials," Science **303**, 1494–1496 (2004).

[15] S Zhang, W. J. Fan, B. K. Minhas, A. Frauenglass, K. J. Malloy, and S. R. J. Brueck, "Midinfrared resonant magnetic nanostructures exhibiting a negative permeability," Phys. Rev. Lett. **94**, 037402 (2005).

[16] J. Butet, K. Thyagarajan, and O. J. F. Martin, "Ultrasensitive optical shape characterization of gold nanoantennas using second harmonic generation," Nano Lett. **13**, 1787–1792 (2013).

[17] G. A. Etxarri, G. R. Medina, F. L. S. Perez, C. Lopez, L. Chantada, F. Scheffold, J. Aizpurua, N. M. Vesperinas, and J. J. Saenz, "Strong magnetic response of submicron silicon particles in the infrared," Opt. Express **19**, 4815–4826 (2011).

[18] A. I. Kuznetsov, A. E. Miroshnichenko, Y. H. Fu, J. Zhang, and B. Lukyanchuk, "Magnetic light," Sci. Rep. **2**, 492 (2012).

[19] S. Liu, M. B. Sinclair, T. S. Mahony, Y. C. Jun, S. Campione, J. Ginn, D. A. Bender, J. R. Wendt, J. F. Ihlefeld, P. G. Clem, J. B. Wright, and I. Brener, "Optical magnetic mirrors without metals," Optica **1**, 250–256 (2014).

[20] O. Aktsipetrov, P. Elyutin, A. Nikulin, and E. Ostrovskaya, "Size effects in optical second-harmonic generation by metallic nanocrystals and semiconductor quantum dots: the role of quantum chaotic dynamics," Phys. Rev. B **51**, 17591–17599 (1995).

[21] B. Lamprecht, A. Leitner, and F. Aussenegg, "Shg studies of plasmon dephasing in nanoparticles," English, Appl. Phys. B **68**, 419–423 (1999).

[22] H. Tuovinen, M. Kauranen, K. Jefimovs, P. Vahimaa, T. Vallius, J. Turunen, N. V. Tkachenko, and H. Lemmetyinen, "Linear and second-order nonlinear optical properties of arrays of noncentrosymmetric gold nanoparticles," J. Nonlin. Opt. Phys. & Materials **11**, 421–432 (2002).

[23] B. K. Canfield, S. Kujala, K. Jefimovs, J. Turunen, and M. Kauranen, "Linear and nonlinear optical responses influenced by broken symmetry in an array of gold nanoparticles," Opt. Express **12**, 5418–5423 (2004).

[24] S. Kujala, B. Canfield, M. Kauranen, Y. Svirko, and J. Turunen, "Multipole Interference in the Second-Harmonic Optical Radiation from Gold Nanoparticles," Phys. Rev. Lett. **98**, 167403 (2007).

[25] R. Czaplicki, H. Husu, R. Siikanen, J. Makitalo, M. Kauranen, J. Laukkanen, J. Lehtolahti, and M. Kuittinen, "Enhancement of second-harmonic generation from metal nanoparticles by passive elements," Phys. Rev. Lett. **110**, 093902 (2013).

[26] V. K. Valev, J. J. Baumberg, B. D. Clercq, N. Braz, X. Zheng, E. J. Osley, S. Vandendriessche, M. Hojeij, C. Blejean, J. Mertens, C. G. Biris, V. Volskiy, M. Ameloot, Y. Ekinci, G. A. E. Vandenbosch, P. A. Warburton, V. V. Moshchalkov, N. C. Panoiu, and T. Verbiest, "Second-Harmonic Generation from Magnetic Metamaterials," Adv. Mater. **26**, 4074–4081 (2014).

[27] J. Zyss, *Molecular nonlinear optics: materials, physics, and devices*, Vol. 260 (Academic press Boston, 1994).

[28] S. Sioncke, T. Verbiest, and A. Persoons, "Second-order nonlinear optical properties of chiral materials," Materials Science and Engineering: R: Reports **42**, 115 –155 (2003).

[29] M. W. Klein, C. Enkrich, M. Wegener, and S. Linden, "Second-harmonic generation from magnetic metamaterials," Science **313**, 502–504 (2006).

[30] M. W. Klein, M. Wegener, N. Feth, and S. Linden, "Experiments on second- and third-harmonic generation from magnetic metamaterials," Opt. Express **15**, 5238–5247 (2007).

[31] E. Kim, F. Wang, W. Wu, Z. Yu, and Y. Shen, "Nonlinear optical spectroscopy of photonic metamaterials," Phys. Rev. B **78**, 113102 (2008).

[32] C. Ciracì, E. Poutrina, M. Scalora, and D. R. Smith, "Origin of second-harmonic generation enhancement in optical split-ring resonators," Phys. Rev. B **85**, 201403 (2012).

[33] E. Poutrina, D. Huang, Y. Urzhumov, and D. R. Smith, "Nonlinear oscillator metamaterial model: numerical and experimental verification," Opt. Express **19**, 8312–8319 (2011).

[34] A. Rose, D. Huang, and D. R. Smith, "Nonlinear Interference and Unidirectional Wave Mixing in Metamaterials," Phys. Rev. Lett. **110**, 063901 (2013).

[35] G. Dolling, C. Enkrich, M. Wegener, J. F. Zhou, C. M. Soukoulis, and S. Linden, "Cut-wire pairs and plate pairs as magnetic atoms for optical metamaterials," Opt. Lett. **30**, 3198–3200 (2005).

[36] V. M. Shalaev, W. Cai, U. K. Chettiar, H.-K. Yuan, A. K. Sarychev, V. P. Drachev, and A. V. Kildishev, "Negative index of refraction in optical metamaterials," Opt. Lett. **30**, 3356–3358 (2005).

[37] C. M. Studio, "Computer simulation technology," GmbH, Darmstadt, Germany (2009).

[38] T. Heinz, "Nonlinear surface electromagnetic phenomena," in (Elsevier, Oxford, 1991) Chap. Second-Order Nonlinear Optical Effects at Surfaces and Interfaces.

[39] D. Krause, C. W. Teplin, and C. T. Rogers, "Optical surface second harmonic measurements of isotropic thin-film metals: Gold, silver, copper, aluminum, and tantalum," J. Appl. Phys. **96**, 3626–3634 (2004).

[40] F. X. Wang, F. J. Rodriguez, W. M. Albers, R. Ahorinta, J. E. Sipe, and M. Kauranen, "Surface and bulk contributions to the second-order nonlinear optical response of a gold film," Phys. Rev. B **80**, 233402 (2009).

[41] J. D. Jackson, *Classical Electrodynamics*, 3rd ed. (John Wiley, 1999).

[42] L. Cao, N. C. Panoiu, R. D. R. Bhat, and R. M. Osgood, "Surface second-harmonic generation from scattering of surface plasmon polaritons from radially symmetric nanostructures," Phys. Rev. B **79**, 235416 (2009).

[43] M. V. Fedoryuk, "The stationary phase method and pseudodifferential operators," Russian Mathematical Surveys **26**, 65 (1971).

[44] J. I. Dadap, J. Shan, and T. F. Heinz, "Theory of optical second-harmonic generation from a sphere of centrosymmetric material: small-particle limit," J. Opt. Soc. Am. B **21**, 1328 (2004).

[45] K. Eisenthal, "Liquid interfaces probed by second-harmonic and sum-frequency spectroscopy," Chem. Rev **96**, 1343–1360 (1996).

[46] K. B. Eisenthal, "Second harmonic spectroscopy of aqueous nano- and microparticle interfaces," Chemical Reviews **106**, 1462–1477 (2006).

[47] R. M. Corn and D. A. Higgins, "Optical second harmonic generation as a probe of surface chemistry," Chem. Rev. **94**, 107–125 (1994).

Chapter 10

Conclusions and future work

Recent advances and novel concepts in photonics have attracted substantial attention, and the design and the experimental investigation of nonlinear optical metamaterials and devices containing 2D materials is thriving, leading to the development of novel applications. While the immediate factors enabling these new developments in photonics consists of advances in nanofabrication and computational techniques, these new concepts could not be implemented without a deep understanding of the theory of interaction of electromagnetic fields with matter. More specifically, there is a need to gain insight into microscopic optical processes on which the functionalities of photonic devices rely. This insight is largely provided by computational investigations, and although computational electromagnetics has been instrumental in this endeavour, many challenging problems remain unexplored or still await a rigorous analysis.

This thesis answered this need for advanced numerical methods and focussed on the theoretical and computational description of optical interactions in periodic structures. It developed numerical methods to enable the analysis of optical phenomena in microscopic structures and the design and optimisation of photonic devices. In particular, I developed computational methods for modelling nonlinear optical processes, a field of growing interest, in diffraction gratings and other periodic structures, and novel 2D materials believed to lead to ground-breaking developments in photonics in the near future. The numerical methods I proposed here and their computer implementations help in the endeavour to fully understand the intricate optical phenomena in complex periodic grating structures.

The next section will detail my original contributions to the area of computational electromagnetics and outline some directions in which my work could be further extended so as to increase its impact.

10.1 Contributions of this work

The contributions of this thesis to the field of photonics are twofold, the first thrust pertaining to computational electromagnetics for photonic structures, the second class concerning the analysis of nonlinear optical phenomena in nonlinear metasurfaces. Specifically, I have

developed a comprehensive formulation of the RCWA to accurately describe the interaction of light with periodic diffraction grating structures. This formulation incorporates important developments to this numerical method and improves it in several significant aspects. Most importantly, in this thesis, I facilitate the accurate calculation of electromagnetic near-fields in the RCWA by remedying numerical artefacts that largely precluded the use of the RCWA for the analysis of modern plasmonic structures and their nonlinear applications, which are extremely reliant on the accuracy with which the electromagnetic near-field is calculated. In addition, the presented formulation incorporates the correct treatment of oblique 2D-periodic structures, improving the accuracy of their numerical simulation.

These improvements to the RCWA do not only help the advancement of numerical analysis by merely providing their mathematical foundations. Instead, they are made available to researchers in the field of photonics: experimentalists without in-depth knowledge of the RCWA and theorists alike. This was achieved by implementing this advanced method into a commercial software code, OmniSim RCWA, which allows its users to better understand and validate experimental results as well as to design and optimise periodic devices with novel functionalities.

Moreover, I provided in my work a numerical method for the efficient modelling of periodically structured 2D materials. This substantial modification of the RCWA distinguishes itself in three ways from previous work: First, my formulation of the RCWA for 2D materials is not limited to graphene but can incorporate arbitrary 2D materials without making assumptions of their effective thickness. Secondly, these materials can be embedded in a periodic device, which can also comprise periodically patterned bulk-materials. And finally, the formulation presented in this thesis, enables the numerical modelling of nonlinear optical effects in these 2D materials. Importantly, these nonlinear calculations are only possible by means of accurate near-field calculations, which further emphasises the importance of this improvement to the RCWA.

In addition to these contributions to the theory of computational photonics, I provided a comprehensive overview of linear and nonlinear optical properties of 2D materials and numerically investigated their resonant behaviour when incorporated in certain photonic structures. To show the capability of this nonlinear RCWA for 2D materials, I used its software implementation to propose designs of resonant periodically patterned waveguides with spectrally selective enhancement of second- and third-harmonic generation.

Based on a different numerical method for linear diffraction, namely the generalised source method, I developed an efficient numerical method for the investigation of bulk higher-harmonic generation in periodic photonic structures. With this method, I validated experimental results pertaining to nonlinear plasmonic metasurfaces, and thus helped the understanding of novel nonlinear optical concepts, namely large nonlinear optical activity

of achiral origin and resonantly enhanced SHG due to higher-order multipolar excitation. To explain the latter, this thesis also provided an analytical formulation for multipolar radiation from periodic lattices of multipoles.

10.2 Future prospects

The research into the remarkable potential of metamaterials and 2D materials emerging from novel and physical properties related to such structures, such as negative index of refraction and optical cloaking, is growing at an increasing rate. Despite the fact that progress has been made throughout the last decade, this potential is far from being exhausted. The computational methods presented in this thesis are highly suitable tools to facilitate this endeavour, as was demonstrated throughout this work; they delivered valuable insights into the nonlinear optical processes in structured materials at the nanoscale. In addition to their current scope, the nonlinear RCWA and nonlinear GSM have increasing future potential. Thus, this work can be extended and built upon in different ways to meet the future needs of physicists and developers of photonic applications and to drive innovation in the field of photonics.

The most important potential extension of the numerical methods presented in this thesis aims at an increasing complexity of the devices that they can model. The formulation of the nonlinear RCWA and the GSM makes use of the undepleted pump approximation, which is a suitable approximation for most practical scenarios and which facilitates the simulation of a large number of nonlinear processes in photonics heterostructures, including sum- and difference-frequency generation and higher-harmonic generation. However, in order to facilitate the study of nonlinear processes that can not be easily described in this approximation, most importantly the optical Kerr effect as its manifestations in self- and cross-phase modulation effects, a self-consistent iteration of the calculation scheme would be required. Such an iteration will, moreover, increase the validity of the simulation process of very high light intensities, for which the undepleted pump approximation might break down.

Although the setup of many experimental investigations is not exactly represented by the plane wave incidence, which is assumed in the computational methods described throughout this thesis, it is a valid approximation in many practical cases. To overcome this limitation, the nonlinear formulations of the RCWA and GSM can be adapted to allow for a more complex spatio-temporal profile of the incident field, namely non-harmonic time dependence and inhomogeneous wavefront. Both these features can be implemented in these two methods by simple a superposition of multiple incident plane waves with different harmonic time dependence and different wavevector. In order to achieve this supposition in the nonlinear part of the calculations, special care would be required.

In contrast to this general input onto a *periodic* structure, it would be worthwhile to

also consider *isolated* objects for investigation by means of the nonlinear RCWA and GSM. Such isolated structures can practically be modelled by the methods in this thesis, simply by employing a large period between an isolated object and periodic copies thereof. However, it would be beneficial to incorporate the known algorithmic techniques for linear aperiodic structures into the nonlinear formulation of the RCWA to increase its scope of application even further.

Higher complexity of a photonic device manifests itself not only through more complex nonlinear processes therein, but also through a more intricate structural composition. Moreover, to predict and design the functionality of a photonic system, it would be highly beneficial to be able to simulate a continuum of different photonic devices, such as diodes, waveguides, gratings, and detectors, which create, guide, manipulate and finally detect the light therein. Each of these photonic components requires accurate and highly efficient numerical simulation techniques. Moreover, their mutual interactions must be fully incorporated in the suit of numerical methods designed for this device "continuum". The numerical methods presented in this thesis are highly suitable to become part of such a set of computational tools for optical simulation of large-scale photonic applications.

The computational requirements of more complex devices and practical settings can be stated as follows: their accurate simulation substantially increases the computational effort, which is necessary to calculate their electromagnetic fields and therefore characterise their functionality. In respect of the nonlinear RCWA and the nonlinear GSM, additional computational efficiency can be achieved in several ways. As we have seen, the runtime of the RCWA and the GSM depends chiefly on the number of harmonics uses in the calculations. Theoretical approaches to reduce the number of harmonics necessary for obtaining converged results would therefore highly improve the efficiency of these methods. The NVF approach to Fourier factorisation fulfils this need to a wide extent, however, by incorporation of adaptive spatial resolution into the mathematical formulation of the nonlinear RCWA and the GSM, this situation could be further improved. The computational efficiency of the nonlinear GSM can be further improved by taking a different theoretical path. Although the structure of the large liner system of equations, which results from the discretisation in the GSM, is already employed by the iterative solver, this iterative solver exhibits some inefficiency in the case of metals and high-index dielectrics. Employing a different iterative linear solver, or by designing a suitable preconditioner for this system of equations, the efficiency of the nonlinear GSM could be further increased. Apart from these theoretical improvements to the numerical methods introduced in this thesis, they can also benefit greatly from more technical changes to their computer implementation. The implementations of both the OmniSim RCWA and the nonlinear GSM used in the completion of this work are parallelised within the shared memory framework and hence take advantage of

modern multi-core computer systems. By changing their software implementation to employ a distributed memory model, they could also harness the computing power of larger computer clusters and high-performance computing facilities. These changes would allow the simulation of larger, more complex photonic devices and substantially decrease the time required to investigate their functionality or to optimise their design.

The two numerical methods at the heart of this thesis, the nonlinear RCWA and the nonlinear GSM, complement each other, as they allow the simulation of higher-harmonic generation in optically nonlinear 2D materials and in the remaining, bulky parts of a diffraction grating, respectively. An important step for their future development would therefore logically be to combine this functionality into a single algorithmic approach, in order to rigorously calculate harmonic generation from all parts of a complex 2D-3D heterostructure simultaneously by using a single numerical method.

Bringing all these ideas into the realm of reality presents a series of intellectual challenges and it excites me to be part of this great endeavour aimed at eventually overcoming all these obstacles.

Appendix A

Bloch's Theorem for diffraction gratings

In this section a variation of Bloch's theorem for gratings is derived. It is similar, but different from the conventional formulation of Bloch's theorem for photonic crystals [1, 2] due to the more general choice of periodic structure.

Let ϵ_r be a permittivity distribution with arbitrary z-dependence, which is periodic in the transverse x-y-plane with primitive lattice vectors $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$:

$$\epsilon_r(\mathbf{r}) = \epsilon_r(\mathbf{r}_\parallel + \mathbf{e}_z z) = \epsilon_r(\mathbf{r}_\parallel + \mathbf{e}_z z + \mathbf{R}_\parallel) \text{ for all } \mathbf{R} \in \mathscr{R} := \mathscr{R}(\mathbf{\Lambda}_1, \mathbf{\Lambda}_2) \tag{A.1}$$

Its 3D Fourier transform with coefficients ϵ_r' can hence be simplified to a Fourier transform in z and a Fourier series in x-y:

$$\epsilon_r(\mathbf{r}_\parallel + \mathbf{e}_z z) = \int_{\mathbb{R}^3} \epsilon_r(\mathbf{h}) d\mathbf{h} = \sum_{\mathbf{G} \in \mathscr{G}_\parallel} \int_{\mathbb{R}} \epsilon_r(\mathbf{G}_\parallel + \mathbf{e}_z h_z) dh_z \tag{A.2}$$

where $\mathscr{G}_\parallel := \{\mathbf{G} \in \mathscr{G}(\mathbf{\Lambda}_1, \mathbf{\Lambda}_2) \text{ with } \mathbf{G}_z = 0\}$ denotes set of reciprocal lattice vectors with vanishing z-component.

Assume that $\mathbf{E}(\mathbf{r}, \omega)$ solves Maxwell's equations Eq. (2.9) for the periodic permittivity $\epsilon_r(\mathbf{r})$ from Eq. (A.1) and consider its 3D Fourier transform

$$\mathbf{E}(\mathbf{r}, \omega) = \int_{\mathbb{R}^3} \mathbf{A}(\mathbf{k}) d\mathbf{k} \tag{A.3}$$

with vectorial amplitude coefficients $\mathbf{A}(\mathbf{k})$. Inserting this and Eq. (A.2) into Maxwell's equations Eq. (2.9d) and Eq. (2.9c) yields for all $\mathbf{r} \in \mathbb{R}^3$

$$0 = \int_{\mathbb{R}^3} \left(\mathbf{k} \times (\mathbf{k} \times \mathbf{A}(\mathbf{k})) + \omega^2 c^{-2} \int_{\mathbb{R}^3} \epsilon(\mathbf{h}) \mathbf{A}(\mathbf{k} - \mathbf{h}) d\mathbf{h} \right) e^{i\mathbf{k}\cdot\mathbf{r}} d\mathbf{k} \tag{A.4}$$

$$0 = \int_{\mathbb{R}^3} \left(\mathbf{k} \times (\mathbf{k} \times \mathbf{A}(\mathbf{k})) + \omega^2 c^{-2} \int_{\mathbb{R}} \sum_{\mathbf{G}_\parallel \in \mathscr{G}_\parallel} \epsilon(\mathbf{G}_\parallel + \mathbf{e}_z h_z) \mathbf{A}(\mathbf{k} - \mathbf{G}_\parallel - \mathbf{e}_z h_z) dh_z \right) e^{i\mathbf{k}\cdot\mathbf{r}} d\mathbf{k}$$

For this integral to vanish for all $\mathbf{r}$, the integrand has to vanish for all $\mathbf{k}$

$$\mathbf{k} \times (\mathbf{k} \times \mathbf{A}(\mathbf{k})) + \omega^2 c^{-2} \int_{\mathbb{R}} \sum_{\mathbf{G}_\parallel \in \mathscr{G}_\parallel} \epsilon(\mathbf{G}_\parallel + \mathbf{e}_z h_z) \mathbf{A}(\mathbf{k} - \mathbf{G}_\parallel - \mathbf{e}_z h_z) dh_z = 0. \quad \text{(A.5)}$$

This implies that for a given $\mathbf{k}$ the value of $\mathbf{A}(\mathbf{k})$ is only related to $\mathbf{A}(\mathbf{k} - \mathbf{G}_\parallel - \mathbf{e}_z h_z)$ for $\mathbf{G}_\parallel \in \mathscr{G}_\parallel$ and all h_z, but is independent of $\mathbf{A}(\mathbf{k} - \mathbf{G}_\parallel - \mathbf{e}_z h_z)$ if $\mathbf{G}_\parallel \notin \mathscr{G}_\parallel$.

Collecting the contributions of $\mathbf{A}(\mathbf{k} - \mathbf{G}_\parallel - \mathbf{e}_z h_z)$ for $\mathbf{G}_\parallel \in \mathscr{G}_\parallel$ and all h_z allows the definition of an independent solution for each $\mathbf{k}_\parallel \in \mathscr{B}$

$$\begin{aligned} \mathbf{E}'_{\mathbf{k}_\parallel}(\mathbf{r}, \omega) &= \int_{\mathbb{R}} \sum_{\mathbf{G}_\parallel \in \mathscr{G}_\parallel} \mathbf{A}(\mathbf{k}_\parallel - \mathbf{G}_\parallel - \mathbf{e}_z h_z) e^{i\left(\mathbf{k}_\parallel - \mathbf{G}_\parallel - \mathbf{e}_z h_z\right) \cdot \mathbf{r}} dh_z \\ &= \sum_{\mathbf{G}_\parallel \in \mathscr{G}_\parallel} \tilde{\mathbf{A}}(\mathbf{k}_\parallel - \mathbf{G}_\parallel, z) e^{-i\mathbf{G}_\parallel \cdot \mathbf{r}_\parallel} e^{i\mathbf{k}_\parallel \cdot \mathbf{r}_\parallel} = \mathbf{u}_{\mathbf{k}_\parallel}(\mathbf{r}_\parallel, z) e^{i\mathbf{k}_\parallel \cdot \mathbf{r}_\parallel} \end{aligned} \quad \text{(A.6)}$$

where the amplitude $\mathbf{u}_{\mathbf{k}_\parallel}(\mathbf{r}_\parallel, z)$ is periodic in $\mathbf{r}_\parallel$ and arbitrary in z, and $e^{i\mathbf{k}_\parallel \cdot \mathbf{r}_\parallel}$ encodes a phase shift in the periodic plane. Finally, an integration of Eq. (A.6) for all vectors in the first Brillouin zone yields the total field $\mathbf{E}$

$$\mathbf{E}(\mathbf{r}, \omega) = \int_{\mathbf{k}_\parallel \in \mathscr{B}} \mathbf{E}'_{\mathbf{k}_\parallel}(\mathbf{r}, \omega).$$

Appendix B

Plane wave expansion by diffraction orders in homogeneous regions

The electromagnetic fields in homogeneous regions with relative permittivity ϵ_r can be easily expanded in a series of plane waves. In the current setting, these homogeneous regions are the cover and substrate, as well as any homogeneous bulk layers in the grating. Due to the periodicity of the remaining grating structure, the EM fields are pseudo-periodic also in homogeneous regions. Therefore only the discrete diffraction orders contribute to the total EM field.

The two different cases of homogeneous layers have already been mentioned: A homogeneous layer inside the grating structure $z = 0, \ldots, h$ is defined by its top and bottom coordinates z^+ and z^-, respectively. By contrast, if the homogeneous layer is the cover (or the substrate), the bottom (top) coordinate is given directly by z^- (z^+)). The top (bottom) coordinate of the cover (substrate) is formally given by $+\infty$ ($-\infty$), which is not suitable for calculations. Therefore $z^+ = z^-$ is chosen here for cover and substrate.

Thus, let $\mathbf{k}_m^\pm = (k_{m,x}, k_{m,y}, \pm k_{m,z})^T = k_m \hat{\mathbf{v}}_m^\pm$ be a given wavevector of diffraction order $m = (m_1, m_2) \in \{-N_1, \ldots, N_1\} \times \{-N_2, \ldots, N_2\}$ in the medium with relative permittivity ϵ_r, with length $k_m = k_0\sqrt{\epsilon_r}$ and directional unit vector $\mathbf{v}_m^\pm$, where $k_{mz} = \sqrt{k_m^2 - k_{mx}^2 - k_{my}^2}$. According to Maxwell's equations 2.9, the directions of the EM fields $\mathbf{E}$ and $\mathbf{H}$ are transverse, i.e. orthogonal to $\mathbf{k}_m$ and to each other. To describe this, one can define the triplet of pairwise orthogonal vectors of unit length $\hat{\mathbf{v}}_m^\pm$, $\hat{\mathbf{p}}_m^\pm$ and $\hat{\mathbf{s}}_m$ with

$$\hat{\mathbf{s}}_m = \frac{\hat{\mathbf{v}}_m^\pm \times \mathbf{e}_z}{\hat{\gamma}_m} = \gamma_m^{-1} \left(k_{my}, -k_{mx}, 0\right)^T \tag{B.1a}$$

$$\hat{\mathbf{p}}_m^\pm = \hat{\mathbf{s}}_m \times \hat{\mathbf{v}}_m^\pm = k_m^{-1}\gamma_m^{-1} \left(\mp k_{mx}k_{mz}, \mp k_{my}k_{mz}, \gamma_m^2\right)^T \tag{B.1b}$$

where $\gamma_m = \sqrt{k_{mx}^2 + k_{my}^2}$ is the transverse length of the wavevector. The total, pseudo-periodic EM fields in this homogeneous region are hence described by the superposition of

the fields of up-/downward propagating diffraction orders and hence given by

$$\mathbf{E}^{\pm}(\mathbf{r}) = \sum_m \left(\hat{\mathbf{s}}_m c_{s,m}^{\pm} + \hat{\mathbf{p}}_m^{\pm} c_{p,m}^{\pm}\right) \exp\left(ik_{mx}x + ik_{my}y \pm ik_{mz}(z - z^{\pm})\right), \tag{B.2a}$$

$$\epsilon_r^{-1/2} Z_0 \mathbf{H}^{\pm}(\mathbf{r}) = \sum_m \left(\hat{\mathbf{p}}_m^{\pm} c_{s,m}^{\pm} + \hat{\mathbf{s}}_m c_{p,m}^{\pm}\right) \exp\left(ik_{mx}x + ik_{my}y \pm ik_{mz}(z - z^{\pm})\right) \tag{B.2b}$$

where $c_{s,m}^{\pm}$ and $c_{p,m}^{\pm}$ denote the transverse electric and transverse magnetic field coefficient, respectively. In this formulation, both the TE coefficients $c_{s,m}^{\pm}$ and the TM coefficients $c_{p,m}^{\pm}$ have the electric field dimension, $\mathrm{V\,m^{-1}}$.

Equation (B.2) can easily be rearranged to agree with the formulation of an RCWA mode given in Eq. (3.53), where the TE and TM parts of the diffraction orders take the role of RCWA modes: $\hat{\mathbf{s}}_m$ and $\hat{\mathbf{p}}_m$ define the Fourier series coefficients of the mode profile, the propagation constant is given by k_{mz}/k_0, and the excitation by $c_{s,m}^{\pm}$ and $c_{p,m}^{\pm}$. Therefore, the fields everywhere in and around the grating can be described by Eq. (3.53).

An additional and important aspect of the plane wave expansion is the magnitude of the energy flow (in $\mathrm{W\,m^{-2}}$) along the principle direction (z) of a diffraction grating [3]:

$$S_m = S_{s,m} + S_{p,m} = \frac{1}{2}\sqrt{\frac{\epsilon_0}{\mu_0}} \mathrm{Re}\left(k_{mz}/k_0\right) \left(|c_{s,m}^{\pm}|^2 + |c_{p,m}^{\pm}|^2\right) \tag{B.3}$$

Appendix C

The discrete Fourier transform and Toeplitz matrices in RCWA and GSM

The two numerical methods in the focus of this thesis, RCWA and GSM, are based on the Fourier series decomposition of the structure and the EM solution of Maxwell's equations. The connection between Fourier series and the discrete Fourier transform (DFT), the important role of Toeplitz matrices, their properties and a fast multiplication of Toeplitz matrices are over viewed here.

C.1 Calculation of Fourier series coefficients by the FFT

The discrete Fourier transform (DFT) and its inverse (iDFT) are defined in 1D for a pair of vectors $Z, r \in \mathbb{C}^M$ [4, Chapter 12]

$$Z_{j+1} = \sum_{k=0}^{M-1} r_{k+1} w^{-jk}, \text{ for } 0 \leq j < M \text{ and} \tag{C.1}$$

$$r_{j+1} = \frac{1}{M} \sum_{j=0}^{M-1} Z_{k+1} w^{jk}, \text{ for } 0 \leq k < M, \text{ respectively,} \tag{C.2}$$

where $w_M = \exp(2\pi i/M)$ is the Mth complex root of -1.

The calculation of the DFT in its original form is computationally not optimal with asymptotic runtime of $\mathcal{O}(M^2)$. Instead, the fast Fourier transform with asymptotic runtime of $\mathcal{O}(M \log M)$ is used, see Ref. [5]. The FFT is used chiefly in two different ways in the RCWA and GSM. The first is in the transverse discretisation of the permittivity, and will be explained now.

Instead of directly evaluating the defining integral of the Fourier series coefficients of e.g. the permittivity Eq. (3.19) by numerical integration, the FFT can be employed after a small modification. The defining integral for the $n_j = -N_j, ..., N_j$ Fourier series coefficient with respect to $v = x$ ($j = 1$) or $v = y$ ($j = 2$) of a scalar function $f(v)$ in RCWA

and GSM is given by

$$[f]_{n_j} = f_{n_j} := \frac{1}{\Lambda_j}\int_0^{\Lambda_1} f(v)\exp(-in_j\frac{2\pi}{\Lambda_1}v)dv \tag{C.3}$$

and is discretised by means of the midpoint rule with $M_j > 2N_j$ points $v_k = \Lambda_j\frac{k+0.5}{M_j}$, $k = 0, ..., M_j - 1$

$$f_{n_j} \approx \frac{1}{M_j}\sum_{k=0}^{M_j-1} f(v_k)\exp(-in_j\frac{2\pi}{\Lambda_j}v_k) = \frac{1}{M_j}\sum_{k=0}^{M_j-1} f(v_k)w_{M_j}^{-n_jk}w_{M_j}^{-0.5n_j} \tag{C.4}$$

Defining $\tilde{n}_j = n_j + N_j \in \{0, \dots, 2N_j\}$ one finds

$$\begin{aligned} Z_{\tilde{n}_j+1} := f_{\tilde{n}_j-N_j} \approx & w_{M_j}^{-0.5\tilde{n}_j}\frac{1}{M_j}\sum_{k=0}^{M_j-1} f(v_k)w_{M_j}^{(k+0.5)N_j}w_{M_j}^{-\tilde{n}_jk} \\ = & w_{M_j}^{-0.5\tilde{n}_j}\frac{1}{M_j}\sum_{k=0}^{M_j-1} r_{k+1}w_{M_j}^{-\tilde{n}_jk} \end{aligned} \tag{C.5}$$

where $r_{k+1} := f(v_k)w_{M_j}^{(k+0.5)N_j}$. Equation (C.5) allows the application of the FFT as it is equivalent to the definition of the DFT Eq. (C.1) up to the factor $w_{M_j}^{-0.5\tilde{n}_j}\frac{1}{M_j}$. The FFT formally returns M_j complex coefficients $Z_{\tilde{n}_j+1}$, $\tilde{n}_j = 0, \dots M_j$. The actual Fourier series coefficients are given by selecting $f_{n_j} = Z_{n_j+N_j+1}$ for $n_j = -N_j, \dots N_j$.

One might ask, what advantage has the choice of a high number of integration points M_j, when in the end, only $2N_j + 1$ coefficients are used. This is answered immediately by the definition of the Fourier series coefficients in terms of the integral Eq. (C.3). The accuracy of all numerically obtained coefficients f_{n_j}, $-N_j \leq n_j \leq N_j$, is determined by the accuracy of the numerical integration Eq. (C.4) and hence by M_j. Therefore it has to be stressed, that the FFT is used simply as the efficient implementation of a numerical integration scheme. This efficiency can be quantified: the computational cost of the FFT to calculate Eq. (C.5) is $\mathcal{O}(M_j \log M_j)$. This is typically much smaller than the cost of the direct evaluation of Eq. (C.4), which is of the order $\mathcal{O}(N_jM_j)$.

This derivation generalises immediately to the 2D integrals, e.g. Eq. (3.19), because it was independent of the coordinates x and y. Analogous considerations allow the Fourier series reconstruction Eq. (3.20) to be performed by the iDFT.

C.2 Fast multiplication of Toeplitz matrices

The second appearance of the DFT in RCWA and GSM is in the fast multiplication of matrices of Fourier series coefficients with vectors, e.g. Eq. (3.26).

This is achieved by making use of the properties of circulant matrices and Toeplitz

matrices [6]. The circulant matrix $\mathsf{C}\{c\}$ of a vector $c = [c_0, \dots c_{N-1}]$ is defined as

$$\mathsf{C}\{c\} = \begin{bmatrix} c_0 & c_{N-1} & c_{N-2} & \dots & c_2 & c_1 \\ c_1 & c_0 & c_{N-1} & \dots & c_3 & c_2 \\ c_2 & c_1 & c_0 & \dots & c_4 & c_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ c_{N-2} & c_{N-3} & c_{N-4} & \dots & c_0 & c_{N-1} \\ c_{N-1} & c_{N-2} & c_{N-3} & \dots & c_1 & c_0 \end{bmatrix} \tag{C.6}$$

The multiplication of $\mathsf{C}\{c\}$ with a vector $x = [x_0, \dots, x_{N-1}]$, $b = \mathsf{C}\{c\}x$ is hence given by the circular convolution of c and x:

$$b_i = (\mathsf{C}\{c\}x)_i = \sum_{j=0}^{N-1} \tilde{c}_{i-j} x_j \;\text{ for } i = 0 \dots, N-1 \tag{C.7}$$

where $\tilde{c}_k = c_{k+N}$ if $k < 0$ and $\tilde{c}_k = c_k$ if $k \geq 0$.

Therefore $C[c]$ diagonalizes under the DFT, i.e.

$$\mathsf{C}\{c\} = \mathsf{W}_N^* \operatorname{diag}(\mathsf{W}_N c) \mathsf{W}_N / N, \tag{C.8}$$

where the entries of W_N are given by $\mathsf{W}_{N;i+1,j+1} = w_N^{ij}$.

Therefore, the multiplication of $\mathsf{C}\{c\}x$ can be rewritten as

$$\mathsf{C}\{c\}x = \mathsf{W}_N^* \operatorname{diag}(\mathsf{W}_N c) \left(\mathsf{W}_N x\right) / N \tag{C.9}$$

where all operations are either multiplication with diagonal matrices ($\mathcal{O}(N)$) or calculations of the DFT ($\mathsf{W}_N x$), which are implemented using the FFT with ($\mathcal{O}(NlogN)$). This constitutes the fast multiplication of circulant matrices.

The matrices of Fourier series coefficients in RCWA and GSM are not circulant, but of Toeplitz type, i.e. for a vector $a = [a_{-N}, \dots, a_0, \dots a_N] \in \mathbb{C}^{2N+1}$ the Toeplitz matrix $\mathsf{T}\{a\} \in \mathbb{C}^{(N+1)\times(N+1)}$ is given by

$$\mathsf{T}\{a\} = \begin{bmatrix} a_0 & a_{-1} & a_{-2} & \dots & a_{-N+1} & a_{-N} \\ a_1 & a_0 & a_{-1} & \dots & a_{-N+2} & a_{-N+1} \\ a_2 & a_1 & a_0 & \dots & a_{-N+3} & a_{-N+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{N-1} & a_{N-2} & a_{N-3} & \dots & a_0 & a_{-1} \\ a_N & a_{N-1} & a_{N-2} & \dots & a_1 & a_0 \end{bmatrix} \tag{C.10}$$

Obviously, each circulant is a Toeplitz matrix, however, a Toeplitz matrix is not necessarily

a circulant matrix. Instead, each Toeplitz matrix can be embedded in a larger circulant matrix of size $(2N+2)\times(2N+2)$

$$\mathsf{C}\{a_{0:N},0,a_{-N:-1}\}=\begin{bmatrix}\mathsf{T}\{a\} & \mathsf{B}\\ \mathsf{B} & \mathsf{T}\{a\}\end{bmatrix} \tag{C.11}$$

where the off-diagonal sub matrix $\mathsf{B}=\mathsf{T}\{a_{1:N},0,a_{-N:-1}\}$ is also of Toeplitz type.

Multiplication of a Toeplitz matrix $\mathsf{T}\{a\}$ with a vector $x\in\mathbb{C}^{N+1}$ can hence be formally written as circulant multiplication

$$\mathsf{T}\{a\}x=\mathsf{C}\{a_{0:N},0,a_{-N:-1}\}\begin{bmatrix}x\\0\end{bmatrix}. \tag{C.12}$$

Therefore Toeplitz matrix-vector multiplication can be performed efficiently, i.e. with asymptotic runtime of $\mathcal{O}(N\log N)$, by means of the FFT as in Eq. (C.9).

One can analogously find for te case of 2D periodicity that the matrices of Fourier series coefficients are block-Toeplitz-Toeplitz matrices. Such a matrix consists of blocks, which are arranged in a Toeplitz fashion and which encode the dependence on one direction of periodicity (e.g. y), and each block itself has the Toeplitz property and encodes the dependence on the other direction of periodicity (x).

Bibliography

[1] J. Joannopoulos, S. Johnson, J. Winn, and R. Meade, *Photonic Crystals: Molding the Flow of Light*, 2nd ed. (Princeton University Press, 2011).

[2] K. Sakoda, *Optical Properties of Photonic Crystals*, 2nd ed., Springer Series in Optical Sciences (Springer, 2004).

[3] J. D. Jackson, *Classical Electrodynamics*, 3rd ed. (John Wiley, 1999).

[4] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes in C*, Vol. 2 (Cambridge University Press, 1992).

[5] *Intel math kernel library. reference manual*, Santa Clara, USA (Intel Corporation, 2009).

[6] R. M. Gray, *Toeplitz and circulant matrices: a review* (now publishers inc, 2006).