

# Scattering from finite structures : an extended Fourier modal method

**Citation for published version (APA):**

Pisarenco, M. (2011). *Scattering from finite structures : an extended Fourier modal method*. [Phd Thesis 1 (Research TU/e / Graduation TU/e), Mathematics and Computer Science]. Technische Universiteit Eindhoven. <https://doi.org/10.6100/IR716275>

**DOI:**

[10.6100/IR716275](https://doi.org/10.6100/IR716275)

**Document status and date:**

Published: 01/01/2011

**Document Version:**

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

**Please check the document version of this publication:**

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.tue.nl/taverne](http://www.tue.nl/taverne)

**Take down policy**

If you believe that this document breaches copyright please contact us at:

[openaccess@tue.nl](mailto:openaccess@tue.nl)

providing details and we will investigate your claim.

# Scattering from Finite Structures

An Extended Fourier Modal Method

Maxim Pisarenco



Scattering from Finite Structures:  
An Extended Fourier Modal Method

Copyright ©2011 by Maxim Pisarenco, Eindhoven, The Netherlands.

All rights are reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior permission of the author.

Cover design by Paul Verspaget and Maxim Pisarenco.

A catalogue record is available from the Eindhoven University of Technology Library

ISBN 978-90-386-2549-2

NUR 919

Subject headings: boundary value problems; differential operators; eigenvalue problems / electromagnetic waves; diffraction / electromagnetic scattering; numerical methods / inverse problems

The work described in this thesis has been carried out under the auspices of

 **ASML** - Veldhoven, The Netherlands.

# Scattering from Finite Structures: An Extended Fourier Modal Method

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de  
Technische Universiteit Eindhoven, op gezag van de  
rector magnificus, prof.dr.ir. C.J. van Duijn, voor een  
commissie aangewezen door het College  
voor Promoties in het openbaar te verdedigen  
op woensdag 7 september 2011 om 16.00 uur

door

Maxim Pisarenco

geboren te Chişinău, Moldavië

Dit proefschrift is goedgekeurd door de promotor:

prof.dr. R.M.M. Mattheij

Copromotor:

dr. J.M.L. Maubach

# Contents

<b>Preface</b>	<b>ix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Lithography . . . . .	1
1.2 Problem description . . . . .	3
1.3 Objectives and main results . . . . .	5
1.4 Outline of the thesis . . . . .	6
<b>2 Mathematical modeling</b>	<b>9</b>
2.1 Maxwell equations . . . . .	9
2.2 Interface conditions . . . . .	11
2.3 Time-harmonic Maxwell equations . . . . .	13
2.4 Problem geometry and incident field . . . . .	15
2.5 Boundary conditions . . . . .	18
2.5.1 Pseudo-periodic boundary condition . . . . .	18
2.5.2 Radiation boundary condition . . . . .	19
2.6 Numerical methods for Maxwell equations . . . . .	20
2.6.1 Finite-difference time-domain method . . . . .	21
2.6.2 Finite element method . . . . .	22
2.6.3 Fourier modal method . . . . .	23
2.6.4 Integral equation methods . . . . .	24
<b>3 Extension of the Fourier modal method for a model problem</b>	<b>27</b>
3.1 Introduction . . . . .	27
3.2 Standard Fourier modal method . . . . .	29
3.3 Artificial periodization with PMLs . . . . .	31
3.4 The contrast-field formulation of the FMM . . . . .	36
3.4.1 Contrast/background decomposition . . . . .	36
3.4.2 Background field solution . . . . .	37
3.4.3 Contrast field solution . . . . .	39
3.5 Numerical results . . . . .	42



---

<b>4</b>	<b>Generalization to arbitrary shapes and illumination</b>	<b>45</b>
4.1	Standard Fourier modal method . . . . .	45
4.1.1	TE-polarization . . . . .	48
4.1.2	TM-polarization . . . . .	50
4.1.3	Conical incidence . . . . .	51
4.2	Aperiodic Fourier modal method . . . . .	54
4.2.1	TE-polarization . . . . .	59
4.2.2	TM-polarization . . . . .	61
4.2.3	Conical incidence . . . . .	63
4.3	Final remarks . . . . .	64
<b>5</b>	<b>Stable solution of the coupled linear systems</b>	<b>67</b>
5.1	Introduction . . . . .	67
5.2	Homogeneous T-matrix and S-matrix algorithms . . . . .	68
5.3	Non-homogeneous S-matrix algorithm . . . . .	71
5.4	Numerical results . . . . .	75
<b>6</b>	<b>Aperiodic Fourier modal method with alternative discretization</b>	<b>79</b>
6.1	Introduction . . . . .	79
6.2	Two equivalent problems . . . . .	81
6.3	Discretization of the background field . . . . .	83
6.4	Recursive linear systems for the vertical problem . . . . .	85
6.5	Non-homogeneous S-matrix algorithm for repeating slices . . . . .	88
6.5.1	Redheffer notation . . . . .	88
6.5.2	Non-homogeneous Redheffer star product . . . . .	89
6.5.3	Global stack description . . . . .	90
6.5.4	Intermediary coefficients . . . . .	92
6.6	Summary of computational costs . . . . .	93
6.7	Far-field recovery . . . . .	94
6.8	Numerical results . . . . .	94
<b>7</b>	<b>Conclusions and suggestions for future work</b>	<b>103</b>
<b>A</b>	<b>Vector calculus identities</b>	<b>107</b>
<b>B</b>	<b>Derivations</b>	<b>109</b>
B.1	Derivation of the convolution term . . . . .	109
B.2	Discretization using the Galerkin approach . . . . .	109
	<b>Bibliography</b>	<b>113</b>
	<b>Index</b>	<b>121</b>

*Contents*

---

vii

Summary	123
Samenvatting	125
Curriculum vitae	127



# Preface

The work described in this thesis has been carried out in the framework of a continuing collaboration between the Centre for Analysis, Scientific computing and Applications (CASA) at Eindhoven University of Technology, and ASML, world leader in the manufacturing of photolithography systems. It was a delight to work on a topic which posed challenging questions of a theoretical nature and in the same time had a direct practical application. Working on this thesis was a new experience for me that I will remember for life and these lines are being written with genuine satisfaction for finishing the manuscript.

I would like to take a moment now to express my gratitude to all people who contributed in one way or another to this thesis. I am indebted to Bob Mattheij for offering me the opportunity to work on a topic which looks as exciting to me now as it did four years ago. I would like to thank him and my copromotor Jos Maubach for their guidance and trust. I consider myself lucky to have Irwan Setija from ASML as a third supervisor. His enthusiasm has been extremely motivating in periods of productive research as well as in moments of stagnation. Our traditional Friday afternoon meetings usually lasted much longer than planned, and were often ended by Irwan's "I am very curious about the new results...". Besides my three supervisors, my defense committee is completed by prof. dr. Gerard Granet, prof. dr. Paul Urbach, prof. dr. Wim Coene and prof. dr. Mark Peletier. I would like to thank them all for the invested time and their valuable comments.

I am grateful to several other people who have had a direct contribution to this work. Nico and Mark, former PhD students who have been part of the CASA-ASML collaboration, performed excellent research which resulted in well-written dissertations I could easily use. Their work facilitated a quick progress in the initial phase of my PhD. Ronald was always around for some useful scientific discussions as well as for programming related questions. Teis Coenen assisted in obtaining the semi-analytical solution of scattering from a dielectric cylinder. Sjoerd Rienstra's provocative question at a local seminar ignited the research presented in Chapter 6.

My colleagues and friends at CASA are responsible for the informal and very friendly working atmosphere. Many thanks to Aga, Ali, Andriy, Antonino, Badr, Bas, Corien, Darcy, David, Davit, Eric, Erwin, Evgeniya, Hans, Iason, Jan-Willem, Kakuba, Kundan, Laura, Lucia, Marco, both Maria's, Mark, Martien, Michiel, Miguel, Mirela, Neda, Nico, Oleg, Patricio, Peter, Rostyslav, Roxana, Shona, Sinatra, Sudhir, Tasnim, Valeriu, Volha, Willem, Yabin, Yves, Zoran. I am also thankful to Enna for all the help on various administrative matters.

I would like to separately thank Kundan, Valeriu, Sudhir and Maria for the enjoyable trips and fruitful scientific collaboration (see [33, 75, 85] for a proof) we have had together at the study groups "Mathematics for Industry" in the Netherlands and Denmark.

I am deeply grateful to my parents for their unlimited love and support. It is thanks to them that I could reach this stage. During my stay in Germany and the Netherlands my sister Irina and her husband Iurie have often been my first resort to turn to for help. Of course, I would like to thank my very cool nephews Andrei and Emil just for the simple fact that they exist and are so much fun to interact with :) Finally, I am thankful to my lovely Inga for her patience, care and love.

*Maxim Pisarenco*

*Eindhoven, July 2011*

# Chapter 1

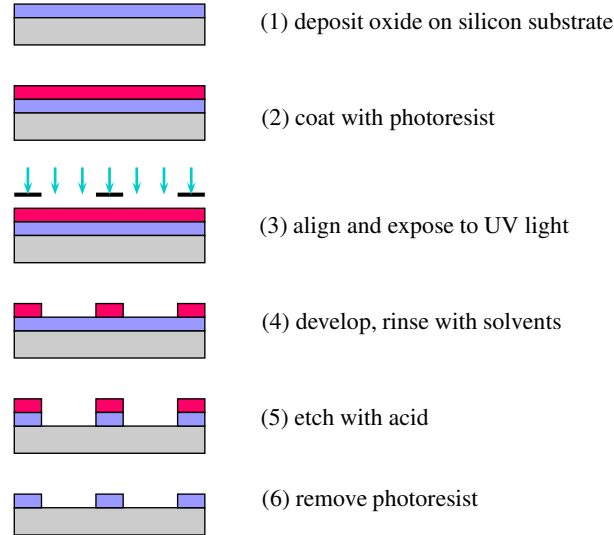
## Introduction

In this chapter we briefly introduce the application which motivates the research documented in this thesis. In particular we describe the fabrication of electronic chips. Then, in the context of inspection and qualification of the quality of printed integrated circuits, a related forward and inverse problem are formulated. We define the objective and list the main results of this thesis. The last section gives a detailed outline of the following chapters.

### 1.1 Lithography

Popular electronic devices of our days (computers, smartphones, TVs, etc.) all share a common ingredient: an *electronic chip* or *integrated circuit*. It is in fact the ingredient. Because the chip essentially operates at the electrical level by either permitting or denying the flow of current through certain connections, it is able to perform considerably faster than any mechanical device (consisting of moving parts). This electronic switching is realized at the lowest level by *transistors*. At a higher level transistors are used to implement the basic logical operations: **not**, **and**, **or**, **xor**, etc. At even a higher level the more complex arithmetic operations can then be performed by a sequence of logical operations. In this manner any programmable algorithm can be reduced to a set of basic operations. The most important property of a chip is that it must be able to regulate its own electrical properties in order to permit or deny the flow of electrical current depending on circumstances. For this reason, it must be made from a *semiconducting material*, such as silicon.

The stages of a lithographic process are shown in Figure 1.1. It starts with a pure silicon



**Figure 1.1:** *Lithographic process.*

ingot, which is cut into thin wafers. Each wafer is then polished and chemically washed to remove scratches and impurities on the surface. To form a thin layer of silicon dioxide on the surface of the wafer (1), the wafer is baked in an oven at 800-1200 °C. Depending upon which process is preferable, a layer of silicon dioxide can also be deposited onto the surface of the wafer using chemical vapor deposition (gas is used to deposit a thin layer of silicon dioxide). Next a coat of photoresist, a light sensitive material, is placed onto the wafer by a process known as spin coating (2). As the photoresist is being placed onto the wafer, the wafer is rotated to create a thin, uniform layer of photoresist. In step (3) ultraviolet light is then shone onto the photoresist-coated wafer through a patterned mask. Masks are larger than the size of the *die* (the area occupied by a single chip on the wafer). For instance, for 157 nm technology the mask is four times the size of the die [38]. If a positive photoresist is used, the photoresist subjected to the light becomes more soluble in the developer solution and is washed away using a solvent, as shown in step (4) of Figure 1.1. On the other hand, if a negative photoresist is used, the photoresist hardens and adheres to the layer of silicon dioxide beneath. When exposed to the developer solution, the unexposed photoresist is washed away, forming a negative image of the mask. The exposed silicon oxide is removed by a chemical etching process (5) while the remaining photoresist protects the unexposed areas with the silicon oxide underneath. Finally the remaining photoresist is removed (6), leaving the intended pattern on the wafer. This process is repeated to form the desired product layer by layer.

The critical step (3) depicted in Figure 1.1 and consisting of alignment and exposure to UV-light is performed by complex *lithography systems*. Figure 1.2 shows such a litho-



**Figure 1.2:** *An ASML lithography machine.*

graphy machine designed and built by ASML<sup>1</sup>. Some of the challenges of step (3) are discussed in the next section.

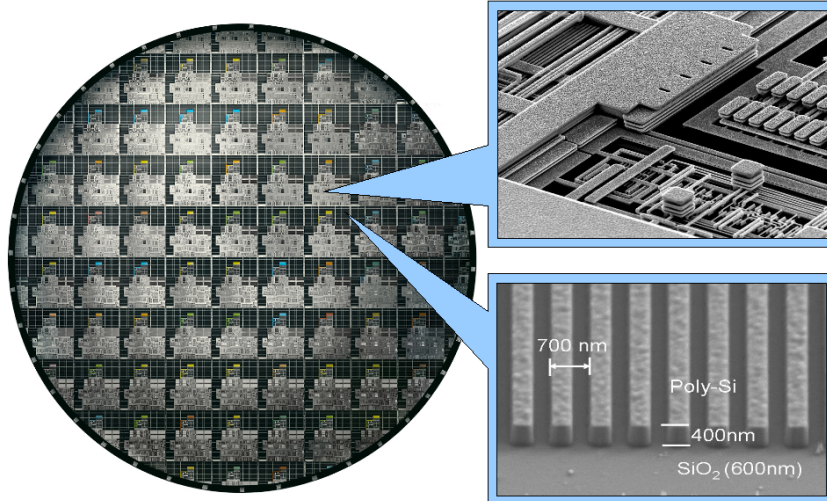
## 1.2 Problem description

The product of a lithographic process is a wafer consisting of an array of electronic chips (Figure 1.3). Figure 1.3 also reveals the presence of *gratings* in the *scribe lane* (the lane between the chips). Unlike electronic chips, gratings have a very regular periodic structure and are much smaller than a chip. Gratings function as metrology targets or markers and serve two purposes: (1) high-precision alignment of consecutively printed layers on the wafer and (2) quality inspection. We explain the quality inspection in more detail. Because of the high complexity, several things can go wrong during a lithographic process: the wafer might be out of focus, the photoresist could be over- or underexposed to UV-light, etc. All these factors influence the quality of the integrated circuit and implicitly the shape of the grating printed in the scribe lane during the same process. Figure 1.4 shows the effect of focus on the shape of the grating lines. The exact grating shape contains information about the quality of the process. The approach of using optical microscopy to determine the shape would fail due to the fact that the size of the grating features (of the order of 100 nm) is smaller than the wavelength of visible

---

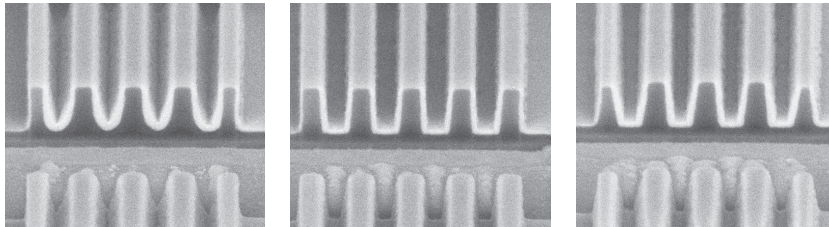
<sup>1</sup>ASML is a company located in Veldhoven, The Netherlands. It is the largest supplier in the world of photolithography systems for the semiconductor industry. The company manufactures machines for the production of integrated circuits, such as RAM and flash memory chips and CPUs.





**Figure 1.3:** The wafer (left) with a zoom-in of the structure of its elements: chip (top right) and grating (bottom right).

light (400...700 nm) and thus beyond the *diffraction limit*. In this range of feature dimensions electron microscopy can be applied. However it has its own weak points: it is slow, expensive and possibly destructive. Particularly, inspection with a scanning electron microscope (SEM) leads to a shrinkage of the grating lines. In order to avoid the drawbacks of a direct measurement (with electron microscopy), an indirect measurement (based on optical metrology) is used. For this purpose, light is shed on the grating, and the intensity of the scattered light is measured by the CCD (charge-coupled device) camera. Figure 1.5 demonstrates the set up for such a measurement. Let  $p$  represent the set of parameters describing the shape of the grating. Two problems can now be formulated.



**Figure 1.4:** Electron microscope image of three gratings. The grating in the middle has been produced when the wafer was in focus. The other two gratings result from a lithographic process with the wafer being out of focus.

**Forward problem (scattering simulation):** Given the grating shape parameters  $p$ , compute the corresponding light intensity  $I(p)$ .

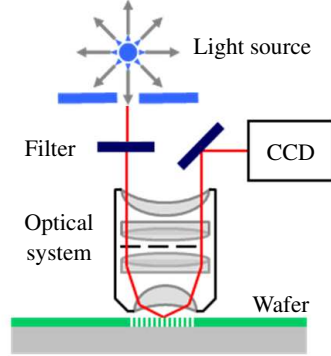
**Inverse problem (profile reconstruction):** Given the measured light intensity  $I_{\text{CCD}}$ , determine the set of grating shape parameters  $p$  that minimizes  $\|I_{\text{CCD}} - I(p)\|$ .

Since light is an electromagnetic wave, a rigorous model of the forward problem is given by the Maxwell equations (see Chapter 2). The inverse problem is also referred to as the reconstruction problem since it is concerned with the reconstruction of the grating shape from the measured light intensity. Often to solve the inverse problem the solution of the forward problem is required. Both problems are non-trivial and extensive research has been done on them. In [88] and [86] the forward and inverse problem respectively are studied in the context of the application addressed in this thesis. An important assumption is made in these studies: the grating is modeled as an infinitely periodic structure. This assumption is accurate enough for gratings which are considerably larger than the size of the illumination spot. However, gratings occupy precious space on the wafer, which could be used for the end product: the electronic chips. Therefore it is desirable to make these structures as small as possible (but large enough to be able to perform reconstruction on them). Another reason for making the gratings smaller is *in-die metrology*, that is metrology on targets placed inside the die (chip area). For small gratings the periodicity assumption introduces a considerable modeling error. This motivates the focus of this thesis: solving the forward problem for finite gratings.

### 1.3 Objectives and main results

Our objective is to *extend the area of application of the Fourier modal method (used in [88] and [86]) from infinitely periodic to finite structures*. The first step in this direction was made by Lalanne and co-workers for waveguide problems [36, 79, 26] where only normal incidence is considered. We build up on this work by applying the Fourier modal method (FMM) to gratings illuminated at arbitrary angles of incidence. Besides the extension of the FMM to finite structures, we are also concerned with *reduction of computational costs in terms of time and memory*. The main results of this thesis are:

- A novel method for simulating scattering from finite structures is formulated: the aperiodic Fourier modal method in contrast field formulation (AFMM-CFF). It inherits the advantages of the standard FMM: simplicity and robustness. In the same time the method is versatile in the sense that it allows an easy switch from periodic boundary conditions to aperiodic ones. In this respect, the AFMM-CFF stands



**Figure 1.5:** *Indirect measurement of the grating profile.*

now in one line with popular numerical methods such as the finite element method (FEM) and the finite-difference time-domain method (FDTD) where both boundary conditions are easily implemented. In comparison to the supercell<sup>2</sup> FMM, the AFMM-CFF achieves the same accuracy with a much smaller number of harmonics. This implies a considerable reduction of time and memory requirements.

- Based on the AFMM-CFF, we develop a method (AFMM-CFF with alternative discretization) which is even faster and uses even less memory. This is achieved by exchanging the discretization directions in the AFMM-CFF, as well as exploiting the *local periodicity* (locally repeating structure) of a finite grating. It is shown in Chapter 6, that the substantial reduction of memory requirements (which can reach factors of 100) is crucial for large-scale problems. We demonstrate that scattering from a large grating with 1024 lines can be relatively easily simulated with the new approach. This is a difficult task for other methods (such as FEM and FDTD), which are generally unable to take advantage of the local periodicity.

## 1.4 Outline of the thesis

**Chapter 2** gives the mathematical description of the physical problem of scattering. The mathematical model consists of the Maxwell equations and constitutive material relations that describe the propagation of electromagnetic waves, as well as conditions that have to be satisfied by the field at material interfaces and domain boundaries. We formulate the equations in time domain and in frequency domain. The description is

<sup>2</sup>In the supercell approach the finite structure is simulated with the standard (periodic) FMM by placing the structure of interest in a large computational domain (or computational cell) such that the interaction with the neighboring cells is minimized.

completed by a definition of an incident field. In the last section of the chapter several popular methods for solving Maxwell's equations are reviewed.

In **Chapter 3** a simple scattering problem is considered: scattering of a TE-polarized plane-wave from a single rectangular line. The principle of the aperiodic FMM in contrast-field formulation (AFMM-CFF) is demonstrated on this "model problem". After the standard FMM is formulated, the idea of artificial periodization with perfectly matched layers (PMLs) is explained. Because the incident field is affected by the PMLs, the problem is reformulated in terms of a contrast field. In this way the incident field is replaced by a virtual source. Results of numerical studies performed with the newly proposed method are presented at the end of the chapter.

The AFMM-CFF is generalized to arbitrary shapes and non-planar illumination in **Chapter 4**. Here both the classical FMM and the AFMM-CFF are presented in their most general formulation. The discretization is formulated in a more elegant manner with the Galerkin approach that is directly applied to the first-order time-harmonic Maxwell equations. The Li rules are applied in the discretization process. The governing equations for the three fundamental cases (TE, TM, conical) are then derived from the discretized equations.

**Chapter 5** focuses on solution strategies for the set of recursive linear systems resulting from the discretization with AFMM-CFF. It is first explained that a straightforward solution approach, the T-matrix algorithm, might encounter instabilities. Then, the homogeneous S-matrix algorithm, used in the classical FMM, is modified and adapted for use with recursive linear systems having non-homogeneous structure. At the end of the chapter, numerical evidence is provided on the stability of the non-homogeneous S-matrix algorithm. For this purpose the problem of scattering by a dielectric cylinder is used, of which a semi-analytical solution is available.

In **Chapter 6** we describe an exchange of spectral and spatial discretization directions in the AFMM-CFF leading to a reduction of computational costs. We start with the observation that computational costs scale cubically with the number of harmonics and linearly with number of slices used in the discretization. This statement suggests that harmonics (being more "expensive") should be used in the shorter direction, while slices (which are "cheaper") can be used in the longer direction. The required modifications of the method are clearly addressed. First the background field is projected on the new basis introduced by the new discretization. Then, an additional reduction of computational costs is obtained by exploiting the local periodicity of the finite grating. The speed-up and memory saving factors are predicted by theoretical estimates and verified by numerical experiments.

Finally, conclusions and suggestions for future investigations are presented in **Chapter 7**.



## Chapter 2

# Mathematical modeling

In this chapter we present the mathematical description of the physical problem of scattering. We first discuss the Maxwell equations and constitutive material relations, which describe the propagation of electromagnetic waves. Next, conditions that have to be satisfied by the field at material interfaces are derived. The mathematical model is completed by a definition of an incident field and a discussion of boundary conditions.

### 2.1 Maxwell equations

Visible light has the physical interpretation of electromagnetic waves with a wavelength between approximately 400 nm and 700 nm. A rigorous model for scattering of light is thus given by the Maxwell equations. In our presentation we follow the classical reference [27]. The macroscopic electromagnetic quantities are related by the *time-dependent Maxwell equations in differential form*:

$$\nabla \times \boldsymbol{\mathcal{E}}(\mathbf{x}, t) + \frac{\partial}{\partial t} \boldsymbol{\mathcal{B}}(\mathbf{x}, t) = \mathbf{0}, \quad (\text{Faraday's law}) \quad (2.1a)$$

$$\nabla \times \boldsymbol{\mathcal{H}}(\mathbf{x}, t) - \frac{\partial}{\partial t} \boldsymbol{\mathcal{D}}(\mathbf{x}, t) = \boldsymbol{\mathcal{J}}, \quad (\text{Ampère's law}) \quad (2.1b)$$

$$\nabla \cdot \boldsymbol{\mathcal{D}}(\mathbf{x}, t) = \rho(\mathbf{x}, t), \quad (\text{Gauss's law for electric fields}) \quad (2.1c)$$

$$\nabla \cdot \boldsymbol{\mathcal{B}}(\mathbf{x}, t) = 0, \quad (\text{Gauss's law for magnetic fields}) \quad (2.1d)$$

where  $\boldsymbol{\mathcal{E}}$  is the *electric field*,  $\boldsymbol{\mathcal{B}}$  is the *magnetic induction*,  $\boldsymbol{\mathcal{H}}$  is the *magnetic field*,  $\boldsymbol{\mathcal{D}}$  is the *electric displacement*. Furthermore,  $\boldsymbol{\mathcal{J}}$  denotes the *electric current density*, and  $\rho$  is the *electric charge density*. The vector  $\mathbf{x} \in \mathbb{R}^3$  contains the space variables and  $t \in \mathbb{R}$  is the time variable. *Faraday's law* gives the effect of a changing magnetic field on the

electric field. Similarly, *Ampère's law* gives the effect of a current and a changing electric field on the magnetic field. *Gauss's law for electric fields* gives the relationship between the electric displacement and the charge density. Finally, *Gauss's law for magnetic fields* expresses the fact that the magnetic field is solenoidal. Implicit in the Maxwell equations is the continuity equation for charge density and current density

$$\nabla \cdot \mathcal{J}(\mathbf{x}, t) = -\frac{\partial}{\partial t} \varrho(\mathbf{x}, t), \quad (2.2)$$

which follows from combining the divergence of (2.1b) with the time derivative of (2.1c) and making use of the vector calculus identity (A.1). Table 2.1 summarizes the quantities in the Maxwell equations and lists their SI-units.

<i>Symbol</i>	<i>Name</i>	<i>SI units</i>	
$\mathcal{E}$	electric field	Volt per meter	$\frac{\text{V}}{\text{m}} = \frac{\text{kg} \cdot \text{m}}{\text{A} \cdot \text{s}^3}$
$\mathcal{H}$	magnetic field	Ampère per meter	$\frac{\text{A}}{\text{m}}$
$\mathcal{D}$	electric displacement	Coulomb per square meter	$\frac{\text{C}}{\text{m}^2} = \frac{\text{A} \cdot \text{s}}{\text{m}^2}$
$\mathcal{B}$	magnetic induction	Tesla	$\text{T} = \frac{\text{kg}}{\text{A} \cdot \text{s}^2}$
$\mathcal{J}$	electric current density	Ampère per square meter	$\frac{\text{A}}{\text{m}^2}$
$\varrho$	electric charge density	Coulomb per cubic meter	$\frac{\text{C}}{\text{m}^3} = \frac{\text{A} \cdot \text{s}}{\text{m}^3}$

**Table 2.1:** *The electromagnetic quantities from Maxwell's equations with their units.*

Maxwell's equations cannot be solved without additional relations which incorporate the material properties. For *linear time-invariant media* the following *constitutive relations* hold:

$$\mathcal{D}(\mathbf{x}, t) = \tilde{\epsilon}(\mathbf{x})\mathcal{E}(\mathbf{x}, t), \quad (2.3a)$$

$$\mathcal{B}(\mathbf{x}, t) = \tilde{\mu}(\mathbf{x})\mathcal{H}(\mathbf{x}, t), \quad (2.3b)$$

$$\mathcal{J}(\mathbf{x}, t) = \tilde{\sigma}(\mathbf{x})\mathcal{E}(\mathbf{x}, t) + \mathcal{J}^{\text{ext}}(\mathbf{x}, t), \quad (2.3c)$$

where  $\tilde{\epsilon}(\mathbf{x})$  is the *electric permittivity*,  $\tilde{\mu}(\mathbf{x})$  is the *magnetic permeability*,  $\sigma$  is the *conductivity* and  $\mathcal{J}^{\text{ext}}(\mathbf{x}, t)$  denotes the *external current*. Table 2.2 summarizes the material properties in the constitutive relations and lists their SI units. In the most general case of anisotropic materials, these quantities are  $3 \times 3$  positive definite tensors. In this thesis we restrict ourselves to isotropic materials, such that the permittivity and permeability are scalar functions of position. They are often expressed in terms of relative quantities:

<i>Symbol</i>	<i>Name</i>	<i>SI units</i>	
$\tilde{\epsilon}$	permittivity	Farad per meter	$\frac{\text{F}}{\text{m}} = \frac{\text{A}^2 \cdot \text{s}^4}{\text{kg} \cdot \text{m}^3}$
$\tilde{\mu}$	permeability	Henry per meter	$\frac{\text{H}}{\text{m}} = \frac{\text{kg} \cdot \text{m}}{\text{A}^2 \cdot \text{s}^2}$
$\tilde{\sigma}$	conductivity	Siemens per meter	$\frac{\text{S}}{\text{m}} = \frac{\text{kg} \cdot \text{m}^3}{\text{A}^2 \cdot \text{s}^3}$

**Table 2.2:** *The material properties from the constitutive relations with their units.*

$$\tilde{\epsilon}(\mathbf{x}) = \epsilon_0 \epsilon^r(\mathbf{x}), \quad \epsilon_0 \approx 8.854 \times 10^{-12} \text{ F/m}, \quad (2.4a)$$

$$\tilde{\mu}(\mathbf{x}) = \mu_0 \mu^r(\mathbf{x}), \quad \mu_0 = 4\pi \times 10^{-7} \text{ H/m}, \quad (2.4b)$$

where  $\epsilon_0$  and  $\mu_0$  represent respectively the free space permittivity and permeability.

## 2.2 Interface conditions

In order to derive the conditions on the fields at interfaces between two materials we need to write the Maxwell equations in integral form. Let  $V$  be a closed volume in space,  $S$  the closed surface bounding it,  $da$  an element of area on the surface and  $\mathbf{n}$  a unit normal to the surface at  $da$  pointing outward from the enclosed volume. The divergence theorem applied to (2.1c) and (2.1d) yields

$$\oint_S \mathcal{D} \cdot \mathbf{n} \, da = \int_V \varrho \, d^3x, \quad (2.5a)$$

$$\oint_S \mathcal{B} \cdot \mathbf{n} \, da = 0. \quad (2.5b)$$

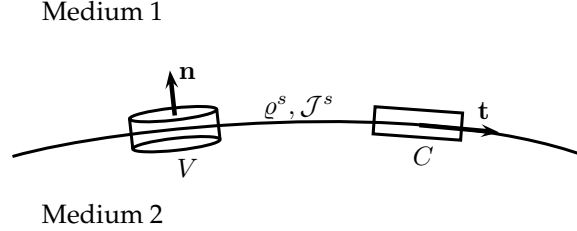
Similarly, let  $C$  be a closed contour in space,  $S'$  the open surface spanning it,  $d\mathbf{l}$  a line element on the contour,  $da$  an element of area on  $S'$ , and  $\mathbf{n}$  a unit normal at  $da$ . The Stokes theorem applied to (2.1a) and (2.1b) yields

$$\oint_C \mathcal{E} \cdot d\mathbf{l} = - \int_{S'} \frac{\partial \mathcal{B}}{\partial t} \cdot \mathbf{n} \, da, \quad (2.5c)$$

$$\oint_C \mathcal{H} \cdot d\mathbf{l} = \int_{S'} \left( \frac{\partial \mathcal{D}}{\partial t} + \mathcal{J} \right) \cdot \mathbf{n} \, da. \quad (2.5d)$$

Equations (2.5) constitute the *time-dependent Maxwell equations in integral form*. We use these to derive the interface conditions. Figure 2.1 depicts the geometrical arrangement used in the derivation. An infinitesimal volume  $V$  in a shape of a pillar box is





**Figure 2.1:** Boundary surface between two media.

placed at the boundary surface between two media. Let  $A$  denote the surface on the interface that also belongs to the volume  $V$ . The integral statements (2.5a) and (2.5b) are applied to the pill box  $V$  with a fixed area  $A$  a vanishing height  $\Delta h \rightarrow 0$ .

$$\int_A (\mathcal{D}_2 - \mathcal{D}_1) \cdot \mathbf{n} \, da = \int_A \varrho^s \, da, \quad (2.6a)$$

$$\int_A (\mathcal{B}_2 - \mathcal{B}_1) \cdot \mathbf{n} \, da = 0. \quad (2.6b)$$

If the charge density  $\varrho$  is singular at the interface such that it determines an idealized *surface charge density*  $\varrho^s$ , then the integral in the right-hand side of (2.6a) is

$$\int_V \varrho \, d^3x = \int_A \varrho^s \, da. \quad (2.7)$$

Because the above relations hold for any area  $A$  on the interface (in other words, the limits of integration along the surface are arbitrary) we have

$$(\mathcal{D}_2 - \mathcal{D}_1) \cdot \mathbf{n} = \varrho^s, \quad (2.8a)$$

$$(\mathcal{B}_2 - \mathcal{B}_1) \cdot \mathbf{n} = 0. \quad (2.8b)$$

Equations (2.8) state that the normal component of the magnetic displacement must be continuous across an interface and the jump in the normal component of the electric displacement is determined by the surface charge density. Now we take an infinitesimal contour  $C$  with a fixed length  $\Delta l$  along the surface and a height  $\Delta h \rightarrow 0$ . The surface  $S'$  spanning  $C$  is oriented so that the normal  $\mathbf{t}$  to  $S'$  is tangent to the interface surface. The integral statements (2.5c) and (2.5d) applied to the contour  $C$  yield

$$\mathbf{n} \times (\mathcal{E}_2 - \mathcal{E}_1) = \mathbf{0}, \quad (2.9a)$$

$$\mathbf{n} \times (\mathcal{H}_2 - \mathcal{H}_1) = \mathcal{J}^s. \quad (2.9b)$$

Equations (2.9) state that the tangential component of the electric field must be continuous across an interface and the jump in the tangential component of the magnetic field is determined by the surface charge density. The terms  $\frac{\partial \mathcal{B}}{\partial t}$ ,  $\frac{\partial \mathcal{D}}{\partial t}$  on the right-hand

side of (2.5c) and (2.5d) vanish because the both are finite at the surface and the area of the loop is zero as the height goes to zero. The idealized surface current density  $\mathcal{J}^s$  corresponds to a current flowing exactly on the interface surface.

$$\int_{\Delta l} \mathcal{J}^s \cdot \mathbf{t} \, dl = \int_{S'} \mathcal{J} \cdot \mathbf{t} \, da. \quad (2.10)$$

In most applications the surface current and the surface charge vanish, i.e.  $\varrho^s = 0$ ,  $\mathcal{J}^s = \mathbf{0}$ .

## 2.3 Time-harmonic Maxwell equations

The time-dependent problem (2.1) can be reduced to the time-harmonic Maxwell system by either using the Fourier transform in time, or by considering propagation of electromagnetic fields with a single frequency. If the electromagnetic waves have an *angular temporal frequency*  $\omega$  (in rad/s), then the electromagnetic quantities are called *time-harmonic* provided they have the following form

$$\mathcal{E}(\mathbf{x}, t) = \Re(e^{i\omega t} \mathbf{e}(\mathbf{x})), \quad (2.11a)$$

$$\mathcal{H}(\mathbf{x}, t) = \Re(e^{i\omega t} \mathbf{h}(\mathbf{x})), \quad (2.11b)$$

$$\mathcal{D}(\mathbf{x}, t) = \Re(e^{i\omega t} \mathbf{d}(\mathbf{x})), \quad (2.11c)$$

$$\mathcal{B}(\mathbf{x}, t) = \Re(e^{i\omega t} \mathbf{b}(\mathbf{x})), \quad (2.11d)$$

$$\mathcal{J}(\mathbf{x}, t) = \Re(e^{i\omega t} \mathbf{j}(\mathbf{x})), \quad (2.11e)$$

$$\varrho(\mathbf{x}, t) = \Re(e^{i\omega t} \rho(\mathbf{x})). \quad (2.11f)$$

Here  $i = \sqrt{-1}$  and  $\Re(\cdot)$  denotes the real part of the expression in parentheses. We use (2.11) in (2.1) to obtain the *time-harmonic Maxwell equations in differential form*

$$\nabla \times \mathbf{e}(\mathbf{x}) = -i\omega \mathbf{b}(\mathbf{x}), \quad (2.12a)$$

$$\nabla \times \mathbf{h}(\mathbf{x}) = i\omega \mathbf{d}(\mathbf{x}) + \mathbf{j}(\mathbf{x}), \quad (2.12b)$$

$$\nabla \cdot \mathbf{d}(\mathbf{x}) = \rho(\mathbf{x}), \quad (2.12c)$$

$$\nabla \cdot \mathbf{b}(\mathbf{x}) = 0. \quad (2.12d)$$

The equations are dependent. This can be seen by taking the divergence of (2.12a) and (2.12b) (making use of (A.1)) to obtain the equations (2.12d) and (2.12c) respectively. The constitutive relations (2.3) do not suffer modifications when written for

time-harmonic quantities:

$$\mathbf{d}(\mathbf{x}) = \tilde{\epsilon}(\mathbf{x})\mathbf{e}(\mathbf{x}), \quad (2.13a)$$

$$\mathbf{b}(\mathbf{x}) = \tilde{\mu}(\mathbf{x})\mathbf{h}(\mathbf{x}), \quad (2.13b)$$

$$\mathbf{j}(\mathbf{x}) = \tilde{\sigma}(\mathbf{x})\mathbf{e}(\mathbf{x}) + \mathbf{j}^{\text{ext}}(\mathbf{x}). \quad (2.13c)$$

Substituting the *constitutive relations for time-harmonic quantities* (2.13) in the time-harmonic Maxwell equations (2.12) yields

$$\nabla \times \mathbf{e}(\mathbf{x}) = -i\omega\mu_0\mu^r\mathbf{h}(\mathbf{x}), \quad (2.14a)$$

$$\nabla \times \mathbf{h}(\mathbf{x}) = (\sigma + i\omega\epsilon_0\epsilon^r)\mathbf{e}(\mathbf{x}) + \mathbf{j}^{\text{ext}}(\mathbf{x}). \quad (2.14b)$$

We introduce the *complex-valued relative permittivity*  $\epsilon = \epsilon^r - i\sigma/(\epsilon_0\omega)$  and assume *source-free* ( $\mathbf{j}^{\text{ext}} = 0$ ) and *non-magnetic materials* ( $\mu^r = 1$ ),

$$\nabla \times \mathbf{e}(\mathbf{x}) = -i\omega\mu_0\mathbf{h}(\mathbf{x}), \quad (2.15a)$$

$$\nabla \times \mathbf{h}(\mathbf{x}) = i\omega\epsilon_0\epsilon\mathbf{e}(\mathbf{x}). \quad (2.15b)$$

In optics it is customary to use the *refractive index*  $n$  to characterize the material. It is related to the permittivity  $\epsilon$  through

$$n = \sqrt{\epsilon} = n' - in'', \quad (2.16)$$

where  $n', n'' \in \mathbb{R}_+$ . Note that the sign of the imaginary part of the refractive index depends on the convention used in the time-harmonic assumption (2.11) (the imaginary part is negative for the  $e^{i\omega t}$  convention and positive for the  $e^{-i\omega t}$  convention). In order to arrive at the final form of the equations, as used in this thesis, the magnetic field is scaled by  $-i\sqrt{\epsilon_0/\mu_0}$  and the *vacuum wavenumber*  $k_0 = \omega\sqrt{\epsilon_0\mu_0}$  is introduced.

$$\nabla \times \mathbf{e}(\mathbf{x}) = -k_0\mathbf{h}(\mathbf{x}), \quad (2.17a)$$

$$\nabla \times \mathbf{h}(\mathbf{x}) = -k_0\epsilon\mathbf{e}(\mathbf{x}). \quad (2.17b)$$

We list the equations in expanded form for future reference:

$$\frac{\partial}{\partial y}e_z - \frac{\partial}{\partial z}e_y = -k_0h_x, \quad (2.18a)$$

$$\frac{\partial}{\partial z}e_x - \frac{\partial}{\partial x}e_z = -k_0h_y, \quad (2.18b)$$

$$\frac{\partial}{\partial x}e_y - \frac{\partial}{\partial y}e_x = -k_0h_z, \quad (2.18c)$$

$$\frac{\partial}{\partial y} h_z - \frac{\partial}{\partial z} h_y = -k_0 \epsilon e_x, \quad (2.18d)$$

$$\frac{\partial}{\partial z} h_x - \frac{\partial}{\partial x} h_z = -k_0 \epsilon e_y, \quad (2.18e)$$

$$\frac{\partial}{\partial x} h_y - \frac{\partial}{\partial y} h_x = -k_0 \epsilon e_z. \quad (2.18f)$$

## 2.4 Problem geometry and incident field

The permittivity  $\epsilon$  in the time-harmonic Maxwell equations (2.18) defines the geometry of the problem. We will consider two-dimensional scatterers, which implies that the permittivity is a function of only two space variables

$$\epsilon = \epsilon(x, z). \quad (2.19)$$

Figure 2.2 shows an example of a grating that is  $y$ -invariant. The grating is supported by a substrate which consists of one or more homogeneous layers. Besides the metrology application described in Section 1.2, this geometry is encountered in many important applications. Examples include broad scientific areas such as submarine detection, geophysical exploration, optical microscopy [42]. In this thesis we consider the general case of a bounded scatterer placed in a *stratified medium*, also referred to as *background multilayer*. The following decomposition can be used for the permittivity

$$\epsilon(x, z) = \epsilon^b(z) + \epsilon^c(x, z), \quad (2.20)$$

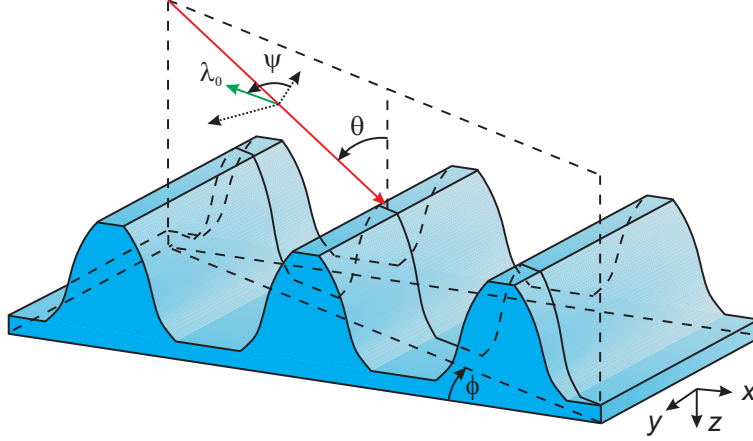
where  $\epsilon^b$  is the *background permittivity* of the stratified medium and  $\epsilon^c$  is the *contrast permittivity* which vanishes outside the bounded scatterer.

The field satisfying the time-harmonic Maxwell equations (2.18) is excited by an incident field. The latter can be a focused beam or uniform illumination. Since any illumination profile can be represented as a superposition of plane waves, we consider the fundamental case of plane wave illumination. It is also assumed that the incident field is time-harmonic and linearly polarized<sup>1</sup>. The orientation of the electromagnetic plane wave with respect to the grating can be defined by three angles: the *polar angle*  $\theta$ , the *azimuthal angle*  $\phi$ , and the *polarization angle*  $\psi$ . These angles are defined in Figure 2.2. Then, the incident electric field is given by

$$\mathbf{e}^{\text{inc}} = \mathbf{a}^e e^{-i\mathbf{k}^{\text{inc}} \cdot \mathbf{x}}, \quad (2.21)$$

---

<sup>1</sup>Linear polarization is a confinement of the electric field vector or magnetic field vector to a given plane along the direction of propagation.



**Figure 2.2:** Angles describing the orientation of wave vector (red) and the amplitude vector (green) of the incident electric field with respect to the grating.

with the wave vector

$$\mathbf{k}^{\text{inc}} = k_0 n_1 \begin{bmatrix} \sin \theta \cos \phi \\ \sin \theta \sin \phi \\ \cos \theta \end{bmatrix}, \quad (2.22)$$

and amplitude vector

$$\begin{aligned} \mathbf{a}^e &= \overbrace{\begin{bmatrix} \cos \psi & \sin \psi & 0 \\ -\sin \psi & \cos \psi & 0 \\ 0 & 0 & 1 \end{bmatrix}}^{\mathbf{R}_\psi} \overbrace{\begin{bmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{bmatrix}}^{\mathbf{R}_\theta} \overbrace{\begin{bmatrix} \cos \phi & \sin \phi & 0 \\ -\sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{bmatrix}}^{\mathbf{R}_\phi} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} \cos \psi \cos \theta \cos \phi - \sin \psi \sin \phi \\ \cos \psi \cos \theta \sin \phi + \sin \psi \cos \phi \\ -\cos \psi \sin \theta \end{bmatrix}. \end{aligned} \quad (2.23)$$

The refractive index  $n_1$  corresponds to the material through which the incident wave is propagating (usually air,  $n_{\text{air}} = 1$ ). The wavelength of the incident plane wave is given by  $\lambda_0 = 2\pi/(n_1 k_0)$ . The matrices  $\mathbf{R}_\psi$ ,  $\mathbf{R}_\theta$ ,  $\mathbf{R}_\phi$  are rotation matrices in the direction of the corresponding angles. Because of the linearity of Maxwell equations, the incident field (being part of the total field) has to satisfy the Maxwell equations in homogeneous space. Equations (2.12c) (without charges,  $\rho = 0$ ) and (2.13a) together with the vector identity (A.2) yield the following constraint on the amplitude vector and the wave vector,

$$\mathbf{a}^e \cdot \mathbf{k}^{\text{inc}} = 0. \quad (2.24)$$

It is easy to check that the vectors defined in (2.22) and (2.23) satisfy this requirement.

The (scaled) incident magnetic field can be derived from the incident electric field using (2.17a)

$$\mathbf{h}^{\text{inc}} \stackrel{\text{(A.3)}}{=} -k_0^{-1} \nabla e^{-i\mathbf{k}^{\text{inc}} \cdot \mathbf{x}} \times \mathbf{a}^e = (ik_0^{-1} \mathbf{k}^{\text{inc}} \times \mathbf{a}^e) e^{-i\mathbf{k}^{\text{inc}} \cdot \mathbf{x}} = \mathbf{a}^h e^{-i\mathbf{k}^{\text{inc}} \cdot \mathbf{x}}, \quad (2.25)$$

with

$$\mathbf{a}^h = in_1 \begin{bmatrix} -\cos \psi \sin \phi - \sin \psi \cos \phi \cos \theta \\ \cos \psi \cos \phi - \sin \psi \sin \phi \cos \theta \\ \sin \psi \cos \phi \sin \theta (\cos \phi + \sin \phi) \end{bmatrix}. \quad (2.26)$$

Let us define the *plane of incidence* as the plane spanned by the wave vector  $\mathbf{k}^{\text{inc}}$  and the  $z$ -axis. In the case of normal incidence ( $\theta = 0$ ) we choose the plane of incidence to coincide with the  $xz$ -plane. We distinguish the following fundamental cases:

- *Planar incidence* corresponds to an azimuthal angle  $\phi = 0$  so that the plane of incidence coincides with the  $xz$ -plane. This case can be further divided into two basic subcases, which can be combined using the superposition principle in order to represent any arbitrary polarization within the planar incidence case.
  - *TE (transverse electric) polarization* corresponds to  $\psi = \frac{\pi}{2}$ , which means that the incident electric field is perpendicular to the plane of incidence and parallel to the  $y$ -axis. Substitution of  $\phi = 0$ ,  $\psi = \frac{\pi}{2}$  in (2.22), (2.23), (2.26) shows that the incident fields have the following form

$$\mathbf{e}^{\text{inc}} = [0, e_y^{\text{inc}}, 0]^T, \quad \mathbf{h}^{\text{inc}} = [h_x^{\text{inc}}, 0, h_z^{\text{inc}}]^T. \quad (2.27)$$

This determines a corresponding form of the resulting fields

$$\mathbf{e} = [0, e_y, 0]^T, \quad \mathbf{h} = [h_x, 0, h_z]^T. \quad (2.28)$$

Because of this form, the Maxwell equations (2.18) can be reduced to a single equation for the  $y$ -component of the electric field,  $e_y$ . Then it is sufficient to impose the incident field by

$$e_y^{\text{inc}} = e^{-ik_0 n_1 (x \sin \theta + z \cos \theta)}. \quad (2.29)$$

- *TM (transverse magnetic) polarization* corresponds to  $\psi = 0$  which means that the incident electric field lies in the plane of incidence. In this case the corresponding incident magnetic field is perpendicular to the plane of incidence and parallel to the grating lines in the  $y$ -direction.

$$\mathbf{e}^{\text{inc}} = [e_x^{\text{inc}}, 0, e_z^{\text{inc}}]^T, \quad \mathbf{h}^{\text{inc}} = [0, h_y^{\text{inc}}, 0]^T. \quad (2.30)$$

This determines a corresponding form of the resulting fields

$$\mathbf{e} = [e_x, 0, e_z]^T, \quad \mathbf{h} = [0, h_y, 0]^T. \quad (2.31)$$

Similarly to the TE case, the Maxwell equations (2.18) can now be reduced to a single equation for the  $y$ -component of the magnetic field,  $h_y$ . Then it is sufficient to impose the incident field by

$$h_y^{\text{inc}} = in_1 e^{-ik_0 n_1 (x \sin \theta + z \cos \theta)}. \quad (2.32)$$

- *Conical incidence* corresponds to an azimuthal angle  $\phi \neq 0$ . This case is more general as it allows for arbitrary angles of incidence and incorporates planar incidence as a special case. Unlike in the planar case, the incident and total fields do not have vanishing components and are respectively of the form

$$\mathbf{e}^{\text{inc}} = [e_x^{\text{inc}}, e_y^{\text{inc}}, e_z^{\text{inc}}]^T, \quad \mathbf{h}^{\text{inc}} = [h_x^{\text{inc}}, h_y^{\text{inc}}, h_z^{\text{inc}}]^T, \quad (2.33)$$

and

$$\mathbf{e} = [e_x, e_y, e_z]^T, \quad \mathbf{h} = [h_x, h_y, h_z]^T. \quad (2.34)$$

## 2.5 Boundary conditions

The Maxwell equations admit infinitely many solutions. In order to restrict the space of solutions and ensure well-posedness of the problem, boundary conditions need to be used. We discuss the pseudo-periodic boundary condition and the radiation condition.

### 2.5.1 Pseudo-periodic boundary condition

The *pseudo-periodic boundary condition* arises from the assumption on infinite periodicity of a structure under consideration. For instance, if we want to simulate scattering from a grating, and the grating is "large enough", we may assume

$$\epsilon(x + n\Lambda, z) = \epsilon(x, z), \quad n \in \mathbb{Z}. \quad (2.35)$$

where  $\Lambda$  is the period of the grating. Periodicity of the permittivity determines the following form of the solution of Maxwell's equations (2.18) [56, p. 8],

$$\mathbf{e}(x, y, z) = e^{-ik_x x} \mathbf{e}^{\text{per}}(x, y, z), \quad (2.36a)$$

$$\mathbf{h}(x, y, z) = e^{-ik_x x} \mathbf{h}^{\text{per}}(x, y, z), \quad (2.36b)$$

where the quantities  $\mathbf{e}^{\text{per}}$ ,  $\mathbf{h}^{\text{per}}$  have the same periodicity as  $\epsilon$ . The scalar  $k_x$  in (2.36) is determined by the requirement that the incoming field (2.21) can also be written in form (2.36). This implies  $k_x = k_x^{\text{inc}}$ . Then the pseudo-periodic boundary condition (also known as *Floquet condition* or *Bloch condition*) is given by

$$\mathbf{e}(x + \Lambda, y, z) = e^{-ik_x^{\text{inc}}\Lambda} \mathbf{e}(x, y, z), \quad (2.37a)$$

$$\mathbf{h}(x + \Lambda, y, z) = e^{-ik_x^{\text{inc}}\Lambda} \mathbf{h}(x, y, z). \quad (2.37b)$$

When the assumption (2.35) on infinite periodicity is valid, this condition allows us to restrict the computational domain to one period only. Note that the use of the pseudo-periodic boundary condition introduces a limitation: because the incident plane wave determines the phase shift we cannot impose an incident field consisting of a superposition of two or more plane waves which do not simultaneously satisfy (2.37).

## 2.5.2 Radiation boundary condition

Physically, a radiation condition ensures that the scattered field  $f$  is propagating away from the obstacle. For a bounded scatterer in homogeneous medium the radiation condition has been formulated mathematically by Sommerfeld [77]. It requires that

$$\lim_{|\mathbf{x}| \rightarrow \infty} |\mathbf{x}|^{\frac{n-1}{2}} \left( \frac{\partial}{\partial |\mathbf{x}|} + ik_0 \right) f = 0, \quad (2.38)$$

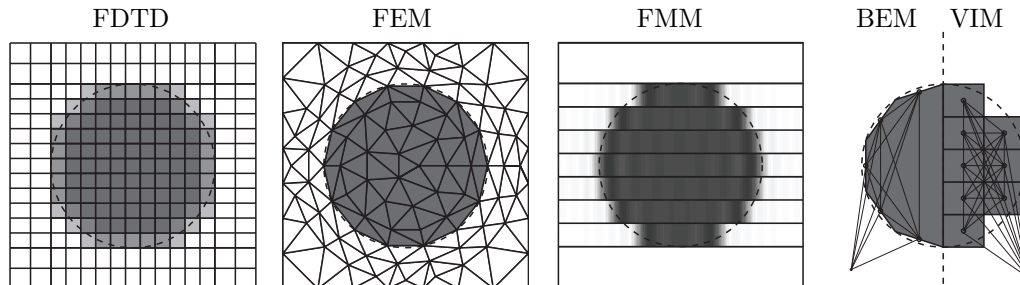
uniformly in all directions. In (2.38)  $n$  is the number of spatial dimensions and  $f$  is a scalar field. The *Sommerfeld radiation condition* can be extended to vector fields. In electromagnetics this extension is known as the *Silver-Müller radiation condition*.

None of the above conditions is suited for scattering from infinitely long interfaces (such as the interfaces in the multilayer stack below the grating). The simple example of scattering from a straight interface where the incident field is a downward propagating plane wave shows that the Sommerfeld radiation condition is not appropriate for such problems: in this case,  $f$  is simply some reflected plane wave and hence satisfies Sommerfeld's radiation condition only in the propagation direction, but no other direction. A natural approach to impose a radiation condition, often used in applications, is to consider the field  $f$  on the truncation boundary and to compute its plane wave expansion. In this expansion we can clearly separate incoming and outgoing waves (the direction of a plane wave is given by its wave vector) and write the field as

$$f = f^{\text{in}} + f^{\text{out}}, \quad (2.39)$$

where  $f^{\text{in}}$  and  $f^{\text{out}}$  represent the fields consisting of incoming and outgoing plane waves





**Figure 2.3:** *The rectangular grid of the FDTD method with sub-pixel smoothing to approximate the cylinder. The triangular grid of the FEM that conforms to the surface of the cylinder. The discrete layers of the FMM, with the permittivity described by a few Fourier modes per layer. The source contributions of the surface elements for the BEM and the polarization densities in the cylinder for the VIM. Image reproduced from [28] with permission.*

respectively. Since the radiation condition admits only outgoing waves we require

$$f^{\text{in}} = 0. \quad (2.40)$$

In the FMM and its extensions described in this thesis, the discretization of the Maxwell equations leads to a plane wave expansion. The radiation condition is then easily imposed on the discrete level by requiring that the coefficients (amplitudes) corresponding to the incoming plane waves vanish.

## 2.6 Numerical methods for Maxwell equations

The number of situations where an analytical solution of Maxwell's equations can be found is very limited. Reference [8] gives a thorough overview of such very special cases. In all other cases numerical methods must be used to get an approximate solution. During the last decades many numerical methods for solving Maxwell equations have been developed. We will discuss several of them: the finite-difference time-domain method (FDTD) [91, 69, 83], the finite-element method (FEM) [50, 93, 51], the Fourier modal method (FMM) [31, 47] and the integral equation methods (IEM), which include the boundary element method (BEM) [68, 42] and the volume integral method (VIM) [7]. A comparison of these methods applied to a particular problem of light diffraction is given in [34].

Figure 2.3 shows the discretization used by the discussed numerical methods for a problem of scattering from a dielectric cylinder. Each of the methods is applied to the Maxwell equations formulated either in time domain (2.1) or frequency domain (2.17).

### 2.6.1 Finite-difference time-domain method

A classical reference on the *finite-difference time-domain* (FDTD) method is [83]. The basic FDTD space grid and time-stepping algorithm trace back to a seminal 1966 paper by Kane Yee [91]. The FDTD method (as the name states) is suitable for solving the Maxwell equations in the time-domain. Spatial discretization uses a structured Cartesian grid (see Figure 2.3). The smooth boundary of the scatterer is approximated by a staircase imposed by the grid. The negative effect of staircasing is sometimes attenuated using subpixel smoothing [16]. The time-dependent equations are discretized using central-difference approximations to the space and time partial derivatives. The resulting finite-difference equations are solved in a leapfrog manner: the electric field vector components in a volume of space are solved at a given instant in time, then the magnetic field vector components in the same spatial volume are solved at the next instant in time. The process is usually repeated until a steady-state is reached.

We demonstrate an FDTD scheme for the case of planar incidence and TE polarization ( $\mathcal{E}_x = \mathcal{E}_z = \mathcal{H}_y = 0$ ). The Maxwell equations (2.1) and the constitutive relations (2.3) reduce to

$$\frac{\partial}{\partial t} \mathcal{H}_x = -\mu^{-1} \frac{\partial}{\partial z} \mathcal{E}_y, \quad (2.41a)$$

$$\frac{\partial}{\partial t} \mathcal{H}_z = \mu^{-1} \frac{\partial}{\partial x} \mathcal{E}_y, \quad (2.41b)$$

$$\frac{\partial}{\partial t} \mathcal{E}_y = \epsilon^{-1} \left( \frac{\partial}{\partial x} \mathcal{H}_z - \frac{\partial}{\partial z} \mathcal{H}_x \right). \quad (2.41c)$$

The *Yee scheme* using a staggered mesh for (2.41) is given by

$$\frac{\mathcal{H}_x|_{i,j}^{n+\frac{1}{2}} - \mathcal{H}_x|_{i,j}^{n-\frac{1}{2}}}{\Delta t} = -\frac{1}{\mu} \frac{\mathcal{E}_y|_{i,j+\frac{1}{2}}^n - \mathcal{E}_y|_{i,j-\frac{1}{2}}^n}{\Delta z}, \quad (2.42a)$$

$$\frac{\mathcal{H}_z|_{i,j}^{n+\frac{1}{2}} - \mathcal{H}_z|_{i,j}^{n-\frac{1}{2}}}{\Delta t} = \frac{1}{\mu} \frac{\mathcal{E}_y|_{i+\frac{1}{2},j}^n - \mathcal{E}_y|_{i-\frac{1}{2},j}^n}{\Delta x}, \quad (2.42b)$$

$$\frac{\mathcal{E}_y|_{i,j}^{n+1} - \mathcal{E}_y|_{i,j}^n}{\Delta t} = \frac{1}{\epsilon} \left( \frac{\mathcal{H}_z|_{i+\frac{1}{2},j}^{n+\frac{1}{2}} - \mathcal{H}_z|_{i-\frac{1}{2},j}^{n+\frac{1}{2}}}{\Delta x} - \frac{\mathcal{H}_x|_{i,j+\frac{1}{2}}^{n+\frac{1}{2}} - \mathcal{H}_x|_{i,j-\frac{1}{2}}^{n+\frac{1}{2}}}{\Delta z} \right), \quad (2.42c)$$

where  $\Delta x$ ,  $\Delta z$  are the mesh sizes and  $\Delta t$  the time step. For a quantity  $\mathcal{F}$  the subscripts indicate the position in the spatial grid and the superscripts indicate the time step,

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

Given the magnetic field at step  $n - \frac{1}{2}$  and the electric field at step  $n$ , the magnetic field at  $n + \frac{1}{2}$  is computed from the update relations (2.42a) and (2.42b). Subsequently the electric field at time step  $n + 1$  can be computed from (2.42c). Because explicit

time-stepping is used in FDTD methods, no linear systems have to be solved. In order to guarantee stability of the computations a limit is imposed on the maximum allowed time step. This requirement is known as the *Courant-Friedrichs-Lewy condition* or *CFL condition* [14].

## 2.6.2 Finite element method

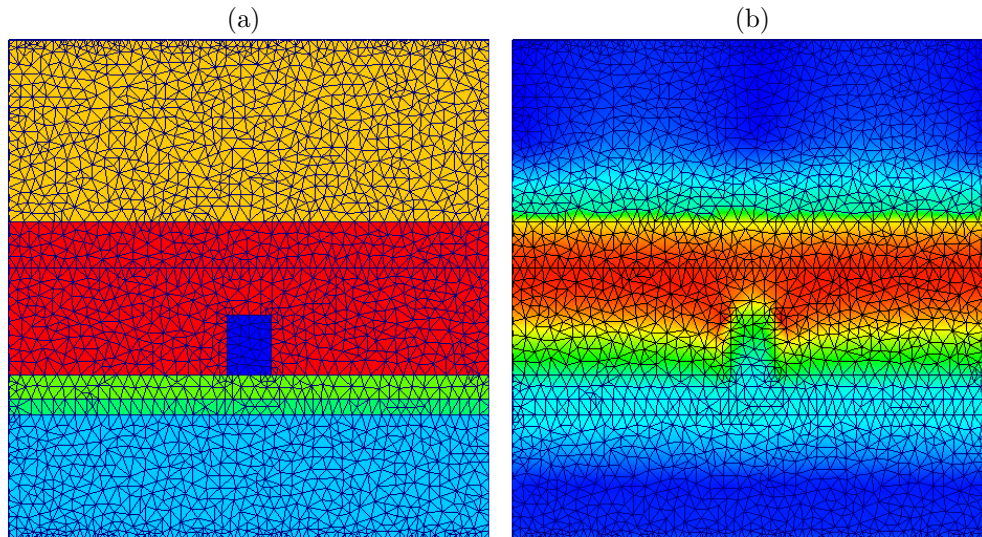
The application of *finite element methods* (FEM) in electromagnetics has been thoroughly described in [51, 80]. A less mathematical description with focus on implementation issues is given in [29]. FEM is applied to Maxwell equations in frequency domain. Typically the following equation is solved:

$$\nabla \times \nabla \times \mathbf{e}(\mathbf{x}) - k_0^2 \epsilon(\mathbf{x}) \mathbf{e}(\mathbf{x}) = \mathbf{0}. \quad (2.43)$$

This equation is referred to as *the double-curl equation* and is obtained by eliminating the magnetic field  $\mathbf{h}$  from (2.17b) using (2.17a). Triangular meshes are used for spatial discretization. They give a better approximation of the smooth shapes than the Cartesian grid of FDTD (see Figure 2.3). The electric field is approximated by a sum of local basis functions. In 1980 Nedelec [53] introduced the vector-valued shape functions called *edge elements*. Contrary to the classical, continuous approximation, the edge elements enforce only the continuity of the tangential component of the electric field. Since the normal component of the electric field is indeed discontinuous at material interfaces, these elements are well-suited for the discretization of the Maxwell equations. Discretization of (2.43) leads to a linear system of the form  $\mathbf{A}\mathbf{x} = \mathbf{b}$ , where  $\mathbf{A}$  is a sparse square matrix. The unknown vector  $\mathbf{x}$  is determined by solving the system using direct [15] or iterative [72] methods.

We have applied FEM to simulate scattering from an infinitely periodic grating. The pseudo-periodicity of the field (Section 2.5.1) implies that a single period of the grating can be considered. Figure 2.4 (a) shows the geometry of the problem with different colors corresponding to different materials. A sample solution for a specific angle of incidence and wavelength is depicted in Figure 2.4 (b).

FEM is a very general numerical method which can be applied to a wide set of partial differential equations (modeling problems in fluid dynamics, solid mechanics, elasticity, acoustics, etc.). The generality of the method makes it also versatile. Techniques such as *adaptive meshing* [23], *multigrid solution* [25, 4], *domain decomposition* [94] have been successfully used when solving electromagnetics problems with FEM.



**Figure 2.4:** Simulation of scattering from an infinitely periodic grating with the FEM: (a) geometry of a single period of the grating (different colors correspond to different materials) and (b) magnitude of the electric field. Both plots show the triangular mesh used for discretization.

### 2.6.3 Fourier modal method

The *Fourier modal method* (FMM) is a numerical solution method of time-harmonic (frequency domain) Maxwell equations for periodic structures. It has originated in the diffractive optics community more than 30 years ago [31]. During this time the method has matured due to improvements to its stability [48, 40] and convergence [41]. Other important contributions to the evolution of the method are the techniques of *adaptive spatial resolution* [22] and *normal vector fields* [65, 66, 78]. Ref. [24] gives a mathematical perspective of the challenges that have been overcome in the FMM and of the open problems still to be addressed.

The discretization used by the FMM is depicted in Figure 2.3. In the vertical direction the domain is divided into layers or slices in which the permittivity is assumed to depend only on the horizontal direction. This introduces a staircase approximation of the geometry, similarly to the Cartesian grid in FDTD. In the horizontal direction Fourier harmonics are used to approximate the fields and the permittivity in each layer. Resulting from the discretization is a set of coupled linear systems which are solved either recursively [48, 40] or are first assembled into a single large linear system of the form  $\mathbf{Ax} = \mathbf{b}$  and then solved with standard routines [52, 46]. Typically the former approach is faster.

The Fourier harmonics constitute a natural basis for wave-like solutions, which makes the

FMM a very popular choice for simulating scattering from infinitely periodic structures. We show in this thesis that if the structure of interest is finite the FMM can be adapted to the new boundary conditions, while keeping the advantages of the periodic FMM.

### 2.6.4 Integral equation methods

The methods described up to now (FDTD, FEM and FMM) use a differential equation and the associated boundary conditions as a starting point for the discretization. A fundamentally different approach is used by the *integral equation methods* (IEM) [43, 74]. The differential equation and the boundary conditions are replaced by an integral equation.

We demonstrate the principle of the IEM on a simple example: scattering of TE-polarized light from a two-dimensional object (such as an infinitely long cylinder) placed in homogeneous medium. The  $y$ -component of the electric field satisfies the equation:

$$\Delta e_y(x, z) + \epsilon(x, z)e_y(x, z) = 0. \quad (2.44)$$

The incident field satisfies a similar equation in homogeneous medium with permittivity  $\epsilon^b$ :

$$\Delta e_y^{\text{inc}}(x, z) + \epsilon^b e_y^{\text{inc}}(x, z) = 0. \quad (2.45)$$

Let  $e_y^{\text{sct}} = e_y - e_y^{\text{inc}}$ . Subtracting (2.45) from (2.44) gives an equation for the scattered field

$$\Delta e_y^{\text{sct}}(x, z) + \epsilon^b e_y^{\text{sct}}(x, z) = -(\epsilon(x, z) - \epsilon^b)e_y(x, z). \quad (2.46)$$

The solution of (2.46) can be written as

$$e_y^{\text{sct}}(x, z) = \int_V G(x - x', z - z') (\epsilon(x', z') - \epsilon^b) e_y(x', z') dx' dz', \quad (2.47)$$

where  $G$  is the *Green's function* and satisfies

$$\Delta G(x, z) + \epsilon^b G(x, z) = -\delta(x, z). \quad (2.48)$$

For the Helmholtz equation in two dimensions the Green's function is given by

$$G(x, z) = \frac{i}{4} H_0^{(1)} \left( \sqrt{\epsilon^b(x^2 + z^2)} \right).$$

Here  $H_0^{(1)}$  is a Hankel function. Finally, the decomposition  $e_y = e_y^{\text{inc}} + e_y^{\text{sct}}$  and relation

(2.47) yield the integral equation

$$e_y(x, z) = e_y^{\text{inc}}(x, z) + \int_V G(x - x', z - z') (\epsilon(x', z') - \epsilon^b) e_y(x', z') dx' dz', \quad (2.49)$$

where the unknown  $e_y$  appears both inside and outside the integral.

In general integral equations arising in electromagnetics are of the form

$$e(\mathbf{x}) = f(\mathbf{x}) + \lambda \int_V K(\mathbf{x}, \mathbf{x}') e(\mathbf{x}') d\mathbf{x}'. \quad (2.50)$$

This is a *Fredholm equation of the second kind*. The function  $K(\mathbf{x}, \mathbf{x}')$  is referred to as *kernel* and  $f(\mathbf{x})$  is a source term. For given  $K(\mathbf{x}, \mathbf{x}')$ ,  $f(\mathbf{x})$  and  $\lambda$ , Equation (2.50) is solved for the unknown  $e(\mathbf{x})$ . Unlike differential equations, integral equations do not require additional (interface or boundary) conditions in order to obtain a unique solution. In fact the interface and boundary conditions are part of the kernel function. A particular kernel is related to a particular geometry and boundary conditions and is not universally valid. The process of solving electromagnetic problems by means of integral equations consists of two steps: {1} formulating the kernel and {2} solving the integral equation.

Methods solving (2.50) are known as *volume integral methods* (VIM). If the volume integral in (2.50) is replaced by a surface integral then we speak of *surface integral methods* (SIM) or *boundary element methods* (BEM). In BEM, instead of the volume of the scatterer, one only has to discretize the surface of the scatterer. If the surface-to-volume ratio is small, then BEM can be considerably more efficient than VIM.



## Chapter 3

# Extension of the Fourier modal method for a model problem

In this chapter we extend the area of application of the Fourier modal method (FMM) from periodic structures to aperiodic ones. This is achieved by placing perfectly matched layers at the lateral sides of the computational domain and reformulating the governing equations in terms of a contrast field which does not contain the incoming field. Due to the reformulation, the homogeneous system of second-order ordinary differential equations from the original FMM becomes non-homogeneous. Its solution is derived analytically and used in the established FMM framework. The technique is demonstrated on an aperiodic model problem of planar scattering of TE-polarized light by a single rectangular line.

### 3.1 Introduction

The Fourier modal method (FMM), also referred to as Rigorous Coupled-Wave Analysis (RCWA), has a well established position in the field of rigorous diffraction modeling. It was first formulated by Moharam and Gaylord in 1981 [45]. Being based on Fourier-mode expansions, the method is inherently suited for (and restricted to) periodic structures such as diffraction gratings. Because harmonic functions constitute a natural basis for representing wave-like solutions few such functions are required to approximate the exact solution with a reasonable accuracy.

The stability and efficiency of the FMM was improved especially due to the *enhanced transmittance matrix approach* for solving the recursive matrix equations [47, 48]. The



convergence problems observed for incident waves with TM-polarization have been overcome by reconsidering the Laurent's rule for the product of truncated Fourier series [35, 21]. Shortly after, these rules have been given a sound mathematical background by Li [41], and usually are referred to as the *Li rules*. The Li rules can be easily applied to 2D-periodic structures with rectangular shapes. For non-rectangular shapes a staircase approximation of the profile in the plane of periodicity had to be used. This inconvenience has been removed by considering separately the tangential and normal components of the field at the interface [65, 66, 78, 70]. Another important improvement was the introduction of the technique of *adaptive spatial resolution* [22]. Due to this technique a faster convergence is achieved by increasing the resolution in space around the material interfaces.

As a consequence of the improvements over the last two decades, nowadays the FMM is a well established method. It is robust and efficient, especially for two-dimensional problems. A recent paper [34] benchmarks the performance of state-of-the-art methods in rigorous diffraction modeling, including the FMM, the finite element method (FEM), the finite difference time-domain method (FDTD) and the volume integral method (VIM).

One important limitation of the FMM is given by the fact that it can only be used for computational problems defined for periodic structures (such as diffraction gratings). This is because the modes used to represent the field are themselves periodic. A straightforward workaround for this limitation is the *supercell approach*: the aperiodic structure is still assumed to be periodic but with a large enough period so that the interaction of neighboring structures is negligible [73].

Lalanne and his co-workers [36, 79, 26] have applied the FMM to waveguide problems. The aperiodicity of the waveguide was dealt with by placing *perfectly matched layers* (PMLs) [5] on the lateral sides of the computational domain. PMLs are introduced in the domain using the mathematical operations of analytic continuation and coordinate transformation. Physically, PMLs represent fictitious absorbing and non-reflecting materials. In this way, artificial periodization is achieved, i.e. the structure of interest is repeated in space, but there is no electromagnetic coupling between neighboring cells. The concepts of perfectly matched layers and artificial periodization are carefully explained in Section 3.3.

The above approach, combining standard FMM with PMLs, is applicable only for the case of normal incidence of the incoming field, which is sufficient for waveguide problems. In this chapter we show that for oblique incidence we need to reformulate the standard FMM such that the incident field is not part of the computed solution. We propose a decomposition of the total field into a *background field* (containing the incident field) and a *contrast field*. The problem is reformulated with the contrast field as the new unknown. The background field solves a corresponding *background problem* which has a standard analytical solution. The main effect of the reformulation is that the homogeneous system



**Figure 3.1:** Geometry of the periodic model problem and division into layers.

of second-order ordinary differential equations becomes non-homogeneous. The solution of this equation is derived in closed form, as required for the FMM algorithm.

The ideas conveyed in this chapter are demonstrated on two model problems: diffraction of TE-polarized light from a binary one-dimensional grating (*periodic model problem*) and from a single line (*aperiodic model problem*). The remainder of the chapter is structured as follows. Section 3.2 briefly describes the standard FMM applied to the periodic model problem. Next, in Section 3.3, the idea of artificial periodization with PMLs is described as a means of solving the aperiodic model problem for normal incidence of the incoming field. Section 3.4 constitutes the core of this chapter and demonstrates the derivation of the contrast-field formulation for the FMM. The new formulation allows for arbitrary angles of incidence in combination with the PMLs. Finally, numerical results are presented in the last section.

## 3.2 Standard Fourier modal method

The structure considered in the periodic model problem is an infinitely periodic binary grating with a period  $\Lambda$  illuminated by a TE-polarized plane wave given by (2.29). The permittivity profile  $\epsilon(x, z)$  is invariant in the  $y$ -direction and is shown in Figure 3.1. The solution of the periodic model problem satisfies the Maxwell equations (2.18), which for TE-polarization can be reduced to a single second-order partial differential equation (PDE) for the  $y$  component of the electric field,

$$\frac{\partial^2}{\partial x^2} e_y(x, z) + \frac{\partial^2}{\partial z^2} e_y(x, z) + k_0^2 \epsilon(x, z) e_y(x, z) = 0. \quad (3.1)$$

The incident field is given by (2.29) and satisfies the pseudo-periodic boundary condition (2.37).

The first step in the FMM is to divide the computational domain into layers such that the permittivity  $\epsilon(x, z)$  is  $z$ -independent in each particular layer. For our periodic model problem this division generates three layers as shown in Figure 3.1. Then the field in

layer  $l$  ( $l = 1, 2, 3$ ) satisfies

$$\frac{\partial^2}{\partial x^2} e_{y,l}(x, z) + \frac{\partial^2}{\partial z^2} e_{y,l}(x, z) + \epsilon_l(x) e_{y,l}(x, z) = 0. \quad (3.2)$$

Note that  $\epsilon_l$ , ( $l = 1, 3$ ) is constant in layers 1 and 3. In this case the solution of (3.2) may be represented by a Rayleigh expansion [56, p. 9]. However, when PMLs are added the Rayleigh expansion is not applicable. For generality we treat these layers in the same way as the middle layer.

The second step in the FMM is to expand the  $x$ -dependent quantities into Fourier modes

$$e_{y,l}(x, z) = \sum_{n=-\infty}^{\infty} s_{n,l}(z) e^{-ik_{xn}x}, \quad (3.3a)$$

$$\epsilon_l(x) = \sum_{n=-\infty}^{\infty} \hat{\epsilon}_{n,l} e^{in\frac{2\pi}{\Lambda}x}, \quad (3.3b)$$

where

$$k_{xn} = k_0 n_1 \sin \theta - n \frac{2\pi}{\Lambda}, \quad n \in \mathbb{Z}.$$

Note that the modes  $e^{-ik_{xn}x}$  satisfy the condition of pseudo-periodicity. Thus, the solution obtained by superposition will necessarily be pseudo-periodic. By substituting the expansions (3.3) in (3.2) and retaining only  $2N + 1$  harmonics in the expansion of the field, we get

$$\begin{aligned} & - \sum_{n=-N}^N k_{xn}^2 s_{n,l}(z) e^{-ik_{xn}x} + \sum_{n=-N}^N \frac{d^2}{dz^2} s_{n,l}(z) e^{-ik_{xn}x} \\ & + \sum_{n=-N}^N \sum_{m=-N}^N \hat{\epsilon}_{n-m,l} s_{m,l}(z) e^{-ik_{xn}x} = 0. \end{aligned} \quad (3.4)$$

The derivation of the third term is detailed in Appendix B.1. Since the functions  $e^{-ik_{xn}x}$  form a basis, the coefficients must vanish.

$$-k_{xn}^2 s_{n,l}(z) + \frac{d^2}{dz^2} s_{n,l}(z) + \sum_{m=-N}^N \hat{\epsilon}_{n-m,l} s_{m,l}(z) = 0, \quad n = -N, \dots, N, \quad (3.5)$$

or in matrix form

$$\frac{d^2}{dz^2} \mathbf{s}_l(z) = k_0^2 \mathbf{A}_l \mathbf{s}_l(z), \quad \text{with } \mathbf{A}_l = \mathbf{K}_x^2 - \mathbf{E}_l, \quad (3.6)$$

where  $\mathbf{K}_x$  is a diagonal matrix with the values  $k_{xn}/k_0$  on its diagonal and  $\mathbf{E}_l$  is a Toeplitz matrix with the  $(n, m)$ -entry equal to  $\epsilon_{n-m,l}$  for  $n, m = -N, \dots, N$ .

Equation (3.6) is a homogeneous second-order ordinary differential equation whose general solution is given by

$$\mathbf{s}_l(z) = \mathbf{s}_l^+(z) + \mathbf{s}_l^-(z) = \mathbf{W}_l(e^{-\mathbf{Q}_l(z-h_{l-1})}\mathbf{c}_l^+ + e^{\mathbf{Q}_l(z-h_l)}\mathbf{c}_l^-), \quad (3.7)$$

where  $h_l$  is the  $z$ -coordinate of the top interface of layer  $l$  (we take  $h_0 = h_1$ ),  $\mathbf{W}_l$  is the matrix of eigenvectors of  $\mathbf{A}_l$  and  $\mathbf{Q}_l$  is a diagonal matrix with square roots of the corresponding eigenvalues on its diagonal.

The general solution (3.7) consists of waves traveling upward,  $\mathbf{s}_l^-(z)$ , and downward,  $\mathbf{s}_l^+(z)$ . In the top and bottom layer the radiation condition is imposed by requiring that there is no incoming field except for the prescribed incident plane wave

$$\mathbf{s}_1^+(h_1) = \mathbf{d}_0 e^{-ik_{z_0}h_1}, \quad (3.8a)$$

$$\mathbf{s}_3^-(h_2) = \mathbf{0}. \quad (3.8b)$$

The vector  $\mathbf{d}_0 \in \mathbb{R}^{2N+1}$  in (3.8a) is an all-zero vector except for entry  $N+1$  which is equal to 1. Conditions (3.8) determine the vectors  $\mathbf{c}_1^+$  and  $\mathbf{c}_3^-$ . The remaining vectors  $\mathbf{c}_l^+$  and  $\mathbf{c}_l^-$  are unknown, and can be determined from the interface conditions between the layers [47]. In the case of the standard FMM the top and bottom layers are homogeneous and Rayleigh expansions of the field can be used. It means that the eigenvalues and eigenvectors for these layers are known in advance.

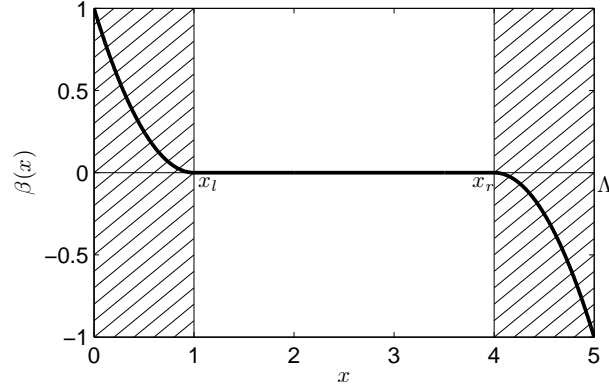
### 3.3 Artificial periodization with PMLs

The goal of this section is to integrate our aperiodic model problem (planar TE diffraction from one line) into the framework of the FMM. To this end we will use the technique of perfectly matched layers which act as absorbing layers and annihilate the effect of the pseudo-periodic boundary conditions. The section starts with a description of the concepts and ideas behind PMLs and ends by explaining the necessity of reformulating the problem in order to allow for arbitrary angles of incidence.

PMLs were first suggested by Berenger [5] as a method of imposing the radiation condition [81] on the boundary of the computational domain in FDTD. According to the formalism proposed by Chew [11, 10], PMLs can be obtained by an analytic continuation of the solution of (3.1) (defined in real coordinates) to a complex contour

$$\tilde{x} = x + i\beta(x), \quad x \in \mathbb{R}. \quad (3.9)$$

The function  $\beta(x)$  is continuous and has a non-zero value only inside the PMLs. For faster convergence also the continuity of higher order derivatives is desirable. Figure 3.2 shows an example of such a function when the PMLs are placed in the domains



**Figure 3.2:** The imaginary part  $\beta(x)$  of the transformation (3.9).

$[0, x_l]$  and  $[x_r, \Lambda]$ . The analytic continuation (3.9) transforms propagating waves into evanescent waves. We may observe the damping effect by evaluating a plane wave on the contour  $\tilde{x}$

$$e^{-i(k_{x0}\tilde{x}+k_{z0}z)} = e^{-i(k_{x0}x+k_{z0}z)} e^{k_{x0}\beta(x)}. \quad (3.10)$$

It is seen that this right-propagating wave (assume  $k_{x0} > 0$ ) is attenuated exponentially in the right PML, where  $\beta(x) < 0$ . The left PML would have the same effect on a left-propagating wave. Since  $k_{x0}$  is in the argument of the real (decaying) exponential, the attenuation in the PML is angle dependent.

The procedure of obtaining a PML requires an analytic continuation from  $\mathbb{R}$  to  $\mathbb{C}$  followed by a coordinate transformation back to  $\mathbb{R}$ . The operations are formally represented by

$$E(x) \xrightarrow{\{1\}} \tilde{E}(\tilde{x}) \xrightarrow{\{2\}} \tilde{E}(x), \text{ with } x \in \mathbb{R}, \tilde{x} \in \mathbb{C}. \quad (3.11)$$

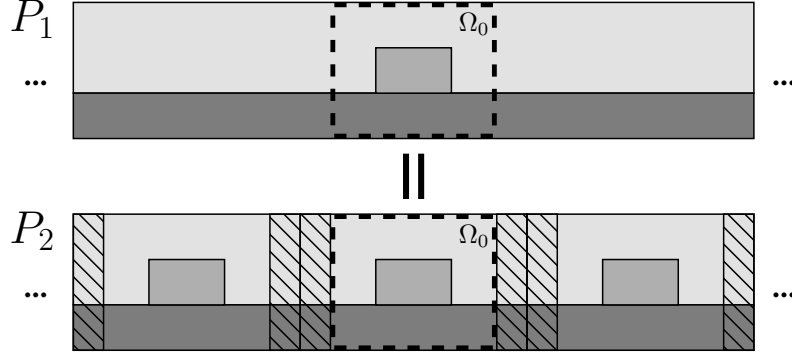
Operation  $\{1\}$  does not formally change the equation but changes its solution by modifying the domain of the space variable  $x$ . Operation  $\{2\}$  is required in order to avoid working in complex coordinates. It is defined as a coordinate transformation

$$\tilde{x} = f(x) = x + i\beta(x), \quad (3.12)$$

applied to the equation in  $\tilde{x}$ . This coordinate transformation eliminates the derivatives with respect to complex variables

$$\frac{\partial}{\partial \tilde{x}} = \frac{dx}{d\tilde{x}} \frac{\partial}{\partial x} = \frac{1}{f'(x)} \frac{\partial}{\partial x}. \quad (3.13)$$

From the above discussion, we conclude that PMLs modify the underlying equations at the continuous level, therefore they can be used in combination with virtually any discretization technique. For the FMM, PMLs are used to make the solution of an



**Figure 3.3:** Problems  $P_1$  (top) and  $P_2$  (bottom) have equal solutions on  $\Omega_0$  (for an ideal non-reflecting PML).

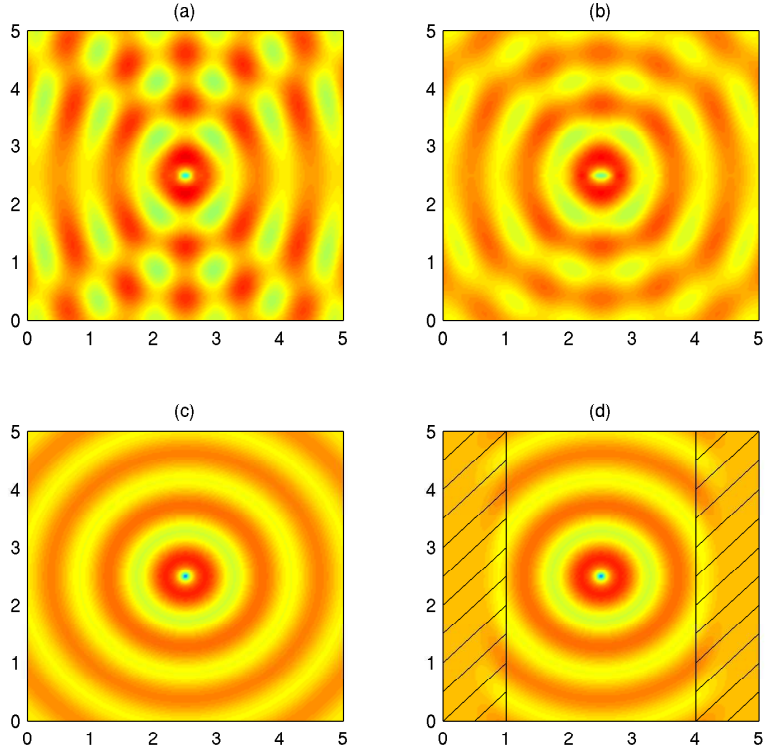
aperiodic problem coincide with the solution of a periodic problem on a subdomain, as explained next.

Suppose we want to solve the aperiodic model problem with normal incidence, that is we want to compute the field scattered by a simple aperiodic structure shown on top of Figure 3.3 (a single rectangular groove infinitely long in the  $y$  direction) when illuminated by a perpendicular plane wave. For this problem, let us refer to it as  $P_1$ , the FMM cannot be used since both the permittivity and field are required to be (pseudo-)periodic functions in order to be represented in terms of Fourier series as in (3.3). However, we can define an equivalent problem  $P_2$  which is artificially periodized with the help of PMLs as shown in Figure 3.3. The problems  $P_1$  and  $P_2$  are equivalent in the sense that (for an ideal PML) their solutions on the domain  $\Omega_0$  coincide. Problem  $P_2$  fits well in the framework of FMM, due to its periodicity.

To solve  $P_2$ , PMLs have to be added. As explained above, PMLs are implemented by {1} an analytic continuation of the solution to a complex contour and {2} a back transformation to the real coordinates. The first step is a formal one as it consists of writing the same partial differential equation in the new variable  $\tilde{x}$  instead of  $x$ . The second step involves the coordinate transformation from  $\tilde{x}$  back to  $x$ . Under this transformation, described by (3.12) and (3.13), Equation (3.2) becomes

$$\frac{1}{f'(x)} \frac{\partial}{\partial x} \left( \frac{1}{f'(x)} \frac{\partial}{\partial x} \tilde{e}_{y,l}(x, z) \right) + \frac{\partial^2}{\partial z^2} \tilde{e}_{y,l}(x, z) + k_0^2 \epsilon_l(x) \tilde{e}_{y,l}(x, z) = 0. \quad (3.14)$$

Note that the permittivity is constant in the PML and is not affected by the transformation. By replacing the field and the permittivity with their corresponding truncated



**Figure 3.4:** Plot of  $\Re(e_y)$  for the radiating line problem solved using various approaches: (a) supercell FMM,  $\Lambda = 10$ , (b) supercell FMM,  $\Lambda = 20$ , (c) exact solution, (d) FMM with PMLs.

Fourier series (as in Section 3.2) the equation can be written in matrix form

$$\frac{d^2}{dz^2} \tilde{\mathbf{s}}_l(z) = k_0^2 \tilde{\mathbf{A}}_l \tilde{\mathbf{s}}_l(z), \quad \tilde{\mathbf{A}}_l = (\mathbf{F}\mathbf{K}_x)^2 - \mathbf{E}_l, \quad (3.15)$$

where  $\mathbf{F}$  is the Toeplitz matrix associated with the Fourier coefficients of  $1/f'(x)$ . Compared to (3.6), the modification introduced by the PML is minor: a "stretching matrix"  $\mathbf{F}$  appears in the computations.

We demonstrate the efficiency of PMLs by considering the problem of a radiating infinitely long line in free space. In a two-dimensional setting the infinitely long line is modeled as a point source. For a source term with vanishing  $x$ - and  $z$ -components, the

following equation holds,

$$\Delta e_y(x, z) + \epsilon^b e_y(x, z) = -\delta(x, z). \quad (3.16)$$

As explained in Section 2.6.4, the solution of the above equation is called a Green's function and is given by

$$e_y(x, z) = \frac{i}{4} H_0^{(1)} \left( \sqrt{\epsilon^b (x^2 + z^2)} \right). \quad (3.17)$$

The solution of (3.16) can also be computed numerically using either the supercell approach or PMLs. The plots (a) and (b) in Figure 3.4 show the supercell solution for domain widths  $\Lambda = 10$  and  $\Lambda = 20$  respectively. Interference caused by the periodic boundary condition decreases for larger sizes of the computational domain. We note that larger computational domains require more harmonics and imply higher computational costs. However, even for  $\Lambda = 20$  the supercell solution is still far from the exact one given by (3.17) and plotted in Figure 3.4 (c). On the other hand, the PML solution shown in Figure 3.4 (d) does not require large computational domains and closely resembles the exact solution in the area between the PMLs. In this example we considered a localized source which is zero in the PMLs. In this case PMLs introduce no complications. For non-localized sources, such as incident fields modeled by plane waves of infinite extent, a reformulation is required. This is explained next.

Since the FMM uses an expansion in pseudo-periodic modes the resulting solution has to be pseudo-periodic. We show that the pseudo-periodicity requirement is only satisfied for normally incident plane waves. We write the total field as a sum of the incident and the scattered field

$$\tilde{e}_y = \tilde{e}_y^{\text{inc}} + \tilde{e}_y^{\text{sct}}.$$

The scattered field (it is an outgoing field) is damped exponentially to "almost zero" at  $x = 0$  and  $x = \Lambda$ . The original incoming field is given by

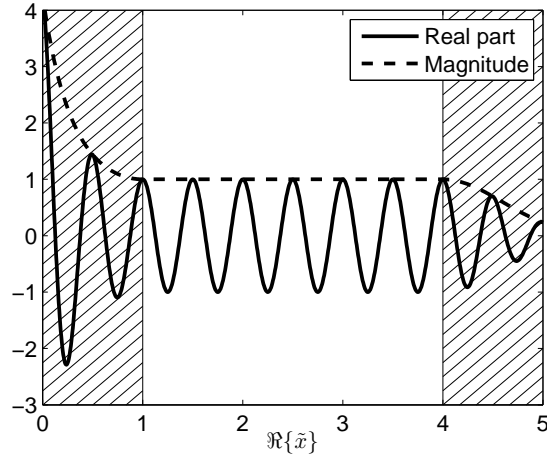
$$e_y^{\text{inc}}(x, z) = e^{-i(k_{x0}x + k_{z0}z)}. \quad (3.18)$$

For normal incidence  $k_{x0} = 0$ , so it is independent of the stretched coordinate  $x$  and is not affected by the PML. Thus, the total field is pseudo-periodic. However, for oblique incidence  $k_{x0} \neq 0$  and the incoming field will be affected by the analytic continuation. We evaluate the incident field on the complex contour  $\tilde{x}$ ,

$$\tilde{e}_y^{\text{inc}}(\tilde{x}, z) = e^{-i(k_{x0}\tilde{x} + k_{z0}z)} = e^{-i(k_{x0}x + k_{z0}z)} e^{k_{x0}\beta(x)}. \quad (3.19)$$

The incident field on the complex contour for  $z = 0$  is plotted in Figure 3.5. Thus, although the scattered field is still damped exponentially to zero at  $x = 0$  and  $x = \Lambda$  and satisfies the pseudo-periodic boundary condition, the incoming field on the complex





**Figure 3.5:** The incident field on the complex contour  $\tilde{x}$  for  $z = 0$ .

contour violates the pseudo-periodicity

$$\tilde{e}_y^{\text{inc}}(f(\Lambda), z) \neq \tilde{e}_y^{\text{inc}}(f(0), z)e^{-ik_{x0}\Lambda}. \quad (3.20)$$

Consequently, also the total field violates this condition and cannot be represented by a superposition of the modes in (3.3a). Therefore, in the next section we remove the part which does not exhibit pseudo-periodicity from the total field and reformulate the problem such that its solution is pseudo-periodic.

## 3.4 The contrast-field formulation of the FMM

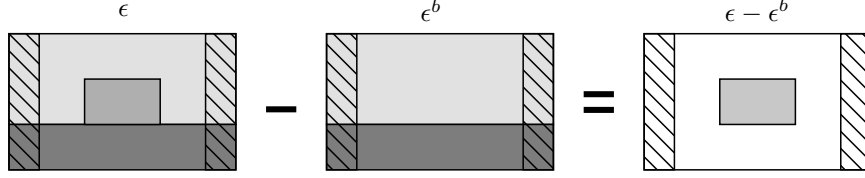
### 3.4.1 Contrast/background decomposition

As shown in the previous section, the presence of PMLs leads to the following form of the governing equation

$$\frac{1}{f'(x)} \frac{\partial}{\partial x} \left( \frac{1}{f'(x)} \frac{\partial}{\partial x} \tilde{e}_y \right) + \frac{\partial^2}{\partial z^2} \tilde{e}_y + k_0^2 \epsilon(x, z) \tilde{e}_y = 0. \quad (3.21)$$

The total field is decomposed into a *contrast field* and a *background field* (this can also be viewed as a decomposition into a periodic part and a non-periodic part)

$$\tilde{e}_y = \tilde{e}_y^c + \tilde{e}_y^b, \quad (3.22)$$



**Figure 3.6:** *Permittivities involved in the source term of (3.24).*

where  $\tilde{e}_y^b$  is chosen to be the field formed in materials defined by  $\epsilon^b(x, z)$

$$\frac{1}{f'(x)} \frac{\partial}{\partial x} \left( \frac{1}{f'(x)} \frac{\partial}{\partial x} \tilde{e}_y^b \right) + \frac{\partial^2}{\partial z^2} \tilde{e}_y^b + k_0^2 \epsilon^b(x, z) \tilde{e}_y^b = 0. \quad (3.23)$$

Subtracting (3.23) from (3.21) yields

$$\frac{1}{f'(x)} \frac{\partial}{\partial x} \left( \frac{1}{f'(x)} \frac{\partial}{\partial x} \tilde{e}_y^c \right) + \frac{\partial^2}{\partial z^2} \tilde{e}_y^c + k_0^2 \epsilon(x, z) \tilde{e}_y^c = -k_0^2 (\epsilon(x, z) - \epsilon^b(x, z)) \tilde{e}_y^b. \quad (3.24)$$

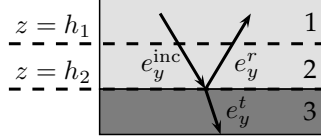
We can still choose  $\epsilon^b$ . However, it should be chosen in such a way that the solution of (3.23) can be computed analytically. Moreover, we want to choose  $\epsilon^b$  such that the right-hand side of (3.24) vanishes in the PML. This is required in order to avoid having a non-periodic source in the PML. If  $\epsilon^b$  is chosen such that it represents the background of  $\epsilon$ , i.e.  $\epsilon$  without the scatterer (rectangular line), then the above mentioned requirements are satisfied; the right-hand side vanishes in the PML, and the background field  $\tilde{e}^b$  can be expressed analytically inside the scatterer. Figure 3.6 shows the permittivities  $\epsilon$ ,  $\epsilon^b$ ,  $\epsilon - \epsilon^b$ , corresponding to the equations for total field (3.21), background field (3.23) and contrast field (3.24).

### 3.4.2 Background field solution

The background field appears on the right-hand side of (3.24). Therefore, before solving (3.24), the solution of (3.23) needs to be found. Since the background field is not pseudo-periodic, we attempt to obtain it analytically and not with the help of FMM. Let us consider the background problem without PMLs

$$\frac{\partial^2}{\partial x^2} e_y^b + \frac{\partial^2}{\partial z^2} e_y^b + k_0^2 \epsilon^b(x, z) e_y^b = 0. \quad (3.25)$$

The solutions  $\tilde{e}_y^b$  and  $e_y^b$  of respectively (3.23) and (3.25) coincide in the physical domain (physical domain = domain - PML region). Since on the right-hand side of (3.24),  $\epsilon - \epsilon^b = 0$  in the PMLs, we need not know the background field in the PML region in



**Figure 3.7:** *The background problem.*

order to solve (3.24).

To solve (3.25), we use knowledge about angles of reflection and refraction. Figure 3.7 shows the representation of the solution in terms of plane waves. We assume  $h_1 = 0$  and  $h_2 = h$ . In layer 2 ( $0 \leq z \leq h$ , see Figure 3.7) the field is written as

$$e_{y,2}^b = e_y^{\text{inc}} + e_y^r = e^{-q_2 z} e^{-ik_{x0}x} + r e^{q_2 z} e^{-ik_{x0}x}. \quad (3.26)$$

In layer 3 ( $z \geq h$ )

$$e_{y,3}^b = e_y^t = t e^{-q_3(z-h)} e^{-ik_{x0}x}, \quad (3.27)$$

where  $q_l = \sqrt{k_{x0}^2 - k_0^2 \epsilon_l^b}$ ,  $l = 2, 3$ . The amplitudes  $r$  and  $t$  corresponding to the reflected and transmitted wave respectively are unknown. They can be determined from the interface conditions, i.e. by matching the fields and their normal derivatives at the interface  $h_2 = h$ ,

$$e_y^{\text{inc}}(x, h) + e_y^r(x, h) = e_y^t(x, h), \quad (3.28a)$$

$$\frac{\partial}{\partial z} e_y^{\text{inc}}(x, h) + \frac{\partial}{\partial z} e_y^r(x, h) = \frac{\partial}{\partial z} e_y^t(x, h). \quad (3.28b)$$

Using the relations (3.26), (3.27) and setting  $b = e^{-q_2 h}$ , we obtain a linear system of equations for  $r$  and  $t$

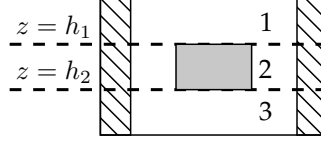
$$r b^{-1} + b = t, \quad (3.29a)$$

$$r q_2 b^{-1} - q_2 b = -t q_3. \quad (3.29b)$$

This system has the solution

$$r = \frac{q_2 - q_3}{q_2 + q_3} b^2, \quad (3.30a)$$

$$t = \frac{2q_2}{q_2 + q_3} b. \quad (3.30b)$$



**Figure 3.8:** The source term in the contrast-field equation.

### 3.4.3 Contrast field solution

The contrast field satisfies Equation (3.24). In layers 1 and 3 the right-hand side vanishes (see Figure 3.8) and the equations are similar to the ones encountered in the standard FMM

$$\frac{1}{f'(x)} \frac{\partial}{\partial x} \left( \frac{1}{f'(x)} \frac{\partial}{\partial x} \tilde{e}_{y,l}^c \right) + \frac{\partial^2}{\partial z^2} \tilde{e}_{y,l}^c + k_0^2 \epsilon_l(x) \tilde{e}_{y,l}^c = 0, \quad l = 1, 3. \quad (3.31)$$

Fourier expansion and truncation yield the system of second-order ordinary differential equations

$$\frac{d^2}{dz^2} \tilde{s}_l^c(z) = k_0^2 \tilde{\mathbf{A}}_l \tilde{s}_l^c(z), \quad l = 1, 3. \quad (3.32)$$

The general solution of this system is given by (3.7). In layer 2 the following equation is solved

$$\frac{1}{f'(x)} \frac{\partial}{\partial x} \left( \frac{1}{f'(x)} \frac{\partial}{\partial x} \tilde{e}_{y,2}^c \right) + \frac{\partial^2}{\partial z^2} \tilde{e}_{y,2}^c + k_0^2 \epsilon_2(x) \tilde{e}_{y,2}^c = -k_0^2 (\epsilon_2(x) - \epsilon_2^b) e_{y,2}^b. \quad (3.33)$$

This equation is non-homogeneous. The following steps are presented in detail since FMM has not been applied to such equations before. We proceed in the usual way by expanding the  $x$ -dependent quantities (consisting of the field and permittivity function) in Fourier modes

$$\tilde{e}_{y,2}^c(x, z) = \sum_{n=-\infty}^{\infty} \tilde{s}_{2,n}^c(z) e^{-ik_{xn}x}, \quad (3.34a)$$

$$e_{y,2}^b(x, z) = e^{-q_2 z} e^{-ik_{x0}x} + r e^{q_2 z} e^{-ik_{x0}x}, \quad (3.34b)$$

$$\epsilon_2(x) = \sum_{n=-\infty}^{\infty} \hat{\epsilon}_{2,n} e^{i \frac{2\pi n}{\Lambda} x}, \quad (3.34c)$$

$$\frac{1}{f'(x)} = \sum_{n=-\infty}^{\infty} \hat{f}_n e^{i \frac{2\pi n}{\Lambda} x}. \quad (3.34d)$$

Substituting (3.34) in (3.33) and truncating the infinite series by keeping the harmonics  $n = -N, \dots, N$  yields

$$\begin{aligned}
& - \sum_{n=-N}^N \sum_{m=-N}^N \left( \hat{f}_{n-m} k_{xm} \sum_{r=-N}^N \hat{f}_{m-r} k_{xr} \tilde{s}_{2,r}^c(z) e^{-ik_{xn}x} \right) \\
& + \sum_{n=-N}^N \frac{d^2}{dz^2} \tilde{s}_{2,n}^c(z) e^{-ik_{xn}x} + k_0^2 \sum_{n=-N}^N \sum_{m=-N}^N \hat{\epsilon}_{2,n-m} \tilde{s}_{2,m}^c(z) e^{-ik_{xn}x} \\
& = -k_0^2 \sum_{n=-N}^N \sum_{m=-N}^N (\hat{\epsilon}_{2,n-m} - \epsilon_2^b \delta_{n-m}) (e^{-q_2 z} + r e^{q_2 z}) \delta_n e^{-ik_{xn}x},
\end{aligned}$$

where  $\delta_n$  is the Kronecker delta ( $\delta_0 = 1, \delta_n = 0, n \in \mathbb{Z} \setminus \{0\}$ ).

Since the functions  $e^{-ik_{xn}x}$  form a basis, their coefficients must satisfy

$$\begin{aligned}
& - \sum_{m=-N}^N \left( \hat{f}_{n-m} k_{xm} \sum_{r=-N}^N \hat{f}_{m-r} k_{xr} \tilde{s}_{2,r}^c(z) \right) + \frac{d^2}{dz^2} \tilde{s}_{2,n}^c(z) + k_0^2 \sum_{m=-N}^N \hat{\epsilon}_{2,n-m} \tilde{s}_{2,m}^c(z) \\
& = -k_0^2 \sum_{m=-N}^N (\hat{\epsilon}_{2,n-m} - \epsilon_2^b \delta_{n-m}) \delta_n (e^{-q_2 z} + r e^{q_2 z}), \quad n = -N, \dots, N.
\end{aligned} \tag{3.35}$$

In matrix form this system of equations can be written as

$$\frac{d^2}{dz^2} \tilde{\mathbf{s}}_2^c(z) = k_0^2 \tilde{\mathbf{A}}_2 \tilde{\mathbf{s}}_2^c(z) + k_0^2 (\epsilon_2^b \mathbf{I} - \mathbf{E}_2) \mathbf{d}_0 (e^{-q_2 z} + r e^{q_2 z}), \tag{3.36}$$

with

$$\tilde{\mathbf{A}}_2 = (\mathbf{F}\mathbf{K}_x)^2 - \mathbf{E}_2.$$

We recall that the vector  $\mathbf{d}_0 \in \mathbb{R}^{2N+1}$  in (3.36) is an all-zero vector except for entry  $N+1$  which is equal to 1. Equation (3.36) is a system of non-homogeneous second order ordinary differential equations (ODEs). The solution vector is of the form

$$\tilde{\mathbf{s}}_2^c = \tilde{\mathbf{s}}_{2,\text{hom}}^c + \tilde{\mathbf{s}}_{2,\text{part}}^c. \tag{3.37}$$

To find the particular solution we use the method of undetermined coefficients applied to systems of equations [82, p. 241]. If the non-homogeneous term contains functions with a finite family of derivatives (e.g. polynomial and trigonometric functions) then the solution may be assumed to be a linear combination of those functions. In our case the particular solution is of the form

$$\tilde{\mathbf{s}}_{2,\text{part}}^c(z) = \mathbf{p} (e^{-q_2 z} + r e^{q_2 z}), \tag{3.38}$$

where  $\mathbf{p} \in \mathbb{R}^{(2N+1)}$  is a vector of coefficients to be determined. We substitute the ansatz (3.38) in Equation (3.36) and get in the end the following linear system which can be

solved for  $\mathbf{p}$

$$(k_0^2 \tilde{\mathbf{A}}_2 - q_2^2 \mathbf{I}) \mathbf{p} = -k_0^2 (\epsilon_2^b \mathbf{I} - \mathbf{E}_2) \mathbf{d}_0. \quad (3.39)$$

Note that in case there is no PML, we have  $\mathbf{F} = \mathbf{I}$  and  $\mathbf{p} = -\mathbf{d}_0$ . The general solution of (3.36) can now be written using (3.37) and (3.38)

$$\tilde{\mathbf{s}}_2^c(z) = \mathbf{W}_2 (e^{-k_0 \mathbf{Q}_2 z} \mathbf{c}_2^+ + e^{k_0 \mathbf{Q}_2 (z-h)} \mathbf{c}_2^-) + \mathbf{p} (e^{-q_2 z} + r e^{q_2 z}). \quad (3.40)$$

The conditions at the layer interface are

$$\tilde{\mathbf{s}}_1^c(0) = \tilde{\mathbf{s}}_2^c(0), \quad (3.41a)$$

$$\frac{1}{k_0} \frac{d}{dz} \tilde{\mathbf{s}}_1^c(0) = \frac{1}{k_0} \frac{d}{dz} \tilde{\mathbf{s}}_2^c(0), \quad (3.41b)$$

$$\tilde{\mathbf{s}}_2^c(h) = \tilde{\mathbf{s}}_3^c(h), \quad (3.41c)$$

$$\frac{1}{k_0} \frac{d}{dz} \tilde{\mathbf{s}}_2^c(h) = \frac{1}{k_0} \frac{d}{dz} \tilde{\mathbf{s}}_3^c(h). \quad (3.41d)$$

Since there is no incident field in the contrast problem, the radiation condition implies

$$\tilde{\mathbf{s}}_1^{c,+}(0) = \mathbf{0}, \quad (3.42a)$$

$$\tilde{\mathbf{s}}_3^{c,-}(h) = \mathbf{0}, \quad (3.42b)$$

where  $\tilde{\mathbf{s}}_1^{c,+}$  and  $\tilde{\mathbf{s}}_3^{c,-}$  represent the fields consisting of respectively downwards and upwards traveling waves (see Equation (3.7)). We set  $\mathbf{V}_l = -\mathbf{W}_l \mathbf{Q}_l$ , and  $\mathbf{X}_2 = e^{-\mathbf{Q}_2 h}$ . By substituting the general solution (3.40) into (3.41), relations for the unknown coefficients three layers are obtained

$$\begin{bmatrix} \mathbf{W}_1 \\ -\mathbf{V}_1 \end{bmatrix} \mathbf{c}_1^- = \begin{bmatrix} \mathbf{W}_2 & \mathbf{W}_2 \mathbf{X}_2 \\ \mathbf{V}_2 & -\mathbf{V}_2 \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \mathbf{c}_2^+ \\ \mathbf{c}_2^- \end{bmatrix} + \begin{bmatrix} \tilde{\mathbf{s}}_{2,\text{part}}^c(0) \\ k_0^{-1} \frac{d}{dz} \tilde{\mathbf{s}}_{2,\text{part}}^c(0) \end{bmatrix}, \quad (3.43)$$

$$\begin{bmatrix} \mathbf{W}_2 \mathbf{X}_2 & \mathbf{W}_2 \\ \mathbf{V}_2 \mathbf{X}_2 & -\mathbf{V}_2 \end{bmatrix} \begin{bmatrix} \mathbf{c}_2^+ \\ \mathbf{c}_2^- \end{bmatrix} + \begin{bmatrix} \tilde{\mathbf{s}}_{2,\text{part}}^c(h) \\ k_0^{-1} \frac{d}{dz} \tilde{\mathbf{s}}_{2,\text{part}}^c(h) \end{bmatrix} = \begin{bmatrix} \mathbf{W}_3 \\ \mathbf{V}_3 \end{bmatrix} \mathbf{c}_3^+, \quad (3.44)$$

where

$$\tilde{\mathbf{s}}_{2,\text{part}}^c(z) = \mathbf{p} (e^{-q_2 z} + r e^{q_2 z}), \quad (3.45)$$

$$\frac{d}{dz} \tilde{\mathbf{s}}_{2,\text{part}}^c(z) = \mathbf{p} (-q_2 e^{-q_2 z} + r q_2 e^{q_2 z}). \quad (3.46)$$

The equations (3.43), (3.44) can be solved for  $\mathbf{c}_1^-$ ,  $\mathbf{c}_2^-$ ,  $\mathbf{c}_2^+$ ,  $\mathbf{c}_3^+$ . Note that the inversion of matrix  $\mathbf{X}_2$  implies growing exponentials and a loss of accuracy due to round-off. This is a general problem encountered by modal methods. Many solutions have been proposed such as the enhanced transmittance matrix approach [48], the S-matrix and R-matrix

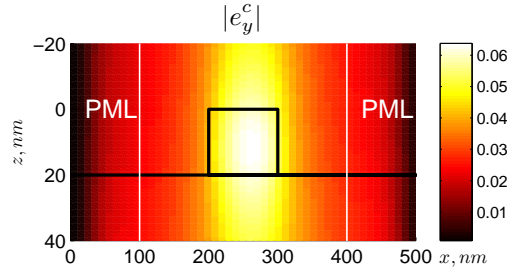


Figure 3.9: The contrast field computed with AFMM-CFF.

algorithms [40]. However, due to the non-homogeneous part, our system has a different structure than in standard FMM and the above algorithms cannot be applied without modifications. We use a *full-matrix approach* [52, 46] in order to guarantee stability.

### 3.5 Numerical results

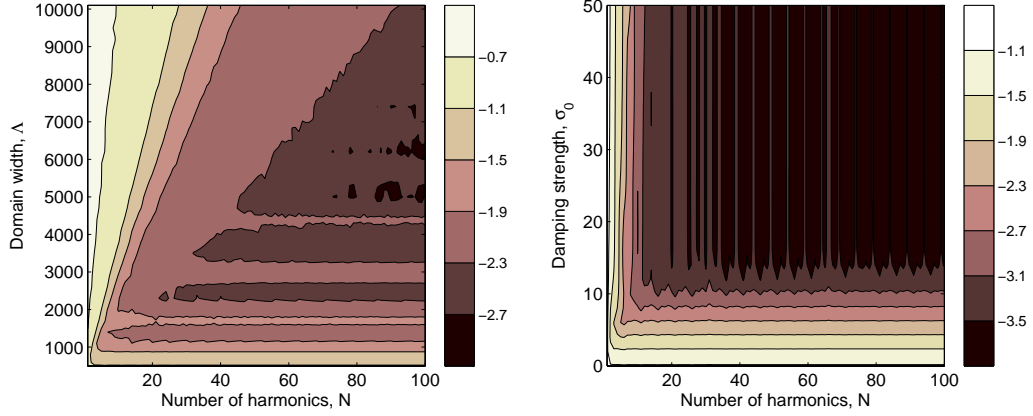
We consider the aperiodic model problem of scattering from an isolated resist line in air with a width of 100 nm and a height of 20 nm illuminated by a plane wave with a wavelength  $\lambda = 628$  nm incident at an angle  $\theta = \pi/6$ . The computational domain has a width  $\Lambda = 500$  nm and the lateral PMLs have a width of 100 nm. The geometry of the problem can be seen in Figure 3.9. The refractive index of air and resist are given by  $n_1 = 1$ ,  $n_3 = 1.5$ .

The contrast-field formulation of the FMM with PMLs is used to solve the problem. We refer to this method as the aperiodic Fourier modal method in contrast-field formulation (AFMM-CFF). For the implementation of the PMLs we need to define the coordinate transformation function which is chosen to be a polynomial of degree  $p$ ,

$$\tilde{x} = f(x) = \begin{cases} x + i\sigma_0|x - x_l|^{(p+1)}/(p+1), & 0 \leq x \leq x_l, \\ x, & x_l < x < x_r, \\ x - i\sigma_0|x - x_r|^{(p+1)}/(p+1), & x_r \leq x \leq \Lambda, \end{cases} \quad (3.47)$$

where  $x_l$  is the endpoint of the left PML,  $x_r$  is the start-point of right PML,  $\sigma_0$  is the damping strength. We chose a quadratic PML ( $p = 2$ ) with a damping strength  $\sigma_0 = 10$ . In the computations also the derivative of the coordinate transformation function is required,

$$\frac{d}{dx}f(x) = \begin{cases} 1 + i\sigma_0|x - x_l|^p, & 0 \leq x \leq x_l, \\ 1, & x_l < x < x_r, \\ 1 - i\sigma_0|x - x_r|^p, & x_r \leq x \leq \Lambda. \end{cases} \quad (3.48)$$



**Figure 3.10:** Logarithmic plots  $\log_{10} \mathcal{E}_1$  (left) and  $\log_{10} \mathcal{E}_2$  (right) of the errors defined in (3.49).

We will first confirm that the PML acts as an absorbing layer. Figure 3.9 shows the contrast field computed with AFMM-CFF. We observe a decay of the field in the PML to "almost zero" at the lateral boundaries, which implies that the effect of the pseudo-periodic boundary condition is negligible. It is clear that the amplitude of the field near the lateral boundary could be used as an indication on the performance of the PML and consequently the accuracy of the numerical solution. However, this approach will only estimate the error due to reflections from the outer boundary and not from the inner boundary of the PML. Note that the solution in the PML is not physically relevant. In order to obtain the solution outside the physical domain, a Green's functions approach may be taken [44].

Next, the convergence behavior of AFMM-CFF and supercell FMM (standard FMM with a large period  $\Lambda$ ) is investigated. For this purpose we define

$$\mathcal{E}_1(N, \Lambda) = \|e_y^{c,N,\Lambda} - e_y^{c,\text{ref}}\|_2, \text{ on } \Omega_s, \quad (3.49a)$$

$$\mathcal{E}_2(N, \sigma_0) = \|\tilde{e}_y^{c,N,\sigma_0} - e_y^{c,\text{ref}}\|_2, \text{ on } \Omega_s, \quad (3.49b)$$

where  $\tilde{e}_y^{c,N,\sigma_0}$  is the numerical solution obtained with AFMM-CFF for  $2N + 1$  harmonics and a damping strength  $\sigma_0$ , while  $e_y^{c,N,\Lambda}$  is the numerical solution obtained with supercell FMM for  $2N + 1$  harmonics and a period  $\Lambda$ . The reference solution is computed using the supercell FMM with  $N = 800$  and  $\Lambda = 15000$ ,  $e_y^{c,\text{ref}} = e_y^{c,800,15000}$ . The Euclidean norm  $\|\cdot\|_2$  is computed on a rectangular domain enclosing the scatterer,  $\Omega_s = [100, 400] \times [-20, 40]$ . Figure 3.10 displays the logarithmic plots of the absolute error for the two methods. Note that since the amplitude of the total field is close to unity, the relative and absolute errors have the same order.



The convergence plots demonstrate that the AFMM-CFF solution converges to the supercell solution. The error of AFMM-CFF has a globally monotonic behavior with respect to  $N$  and  $\sigma_0$ . In other words, increasing either  $N$  or  $\sigma_0$  will not worsen the accuracy of the solution. The supercell FMM has a non-monotonic behavior with respect to  $\Lambda$ . In order to obtain a better solution, increasing  $\Lambda$  would require also taking more harmonics. This behavior is clearly undesirable from the computational point of view.

Also quantitative statements may be made based on Figure 3.10. It indicates that the AFMM-CFF exhibits faster convergence. For instance, an absolute error in the range  $10^{-2.7} \dots 10^{-3}$ , is attained by the supercell FMM for  $N$  around 80, and by the AFMM-CFF for  $N \approx 10$  (the plots have different color scales). Since the methods have cubical complexity with respect to the number of harmonics a speed-up by a factor  $8^3 \approx 500$  can be reached.

## Chapter 4

# Generalization to arbitrary shapes and illumination

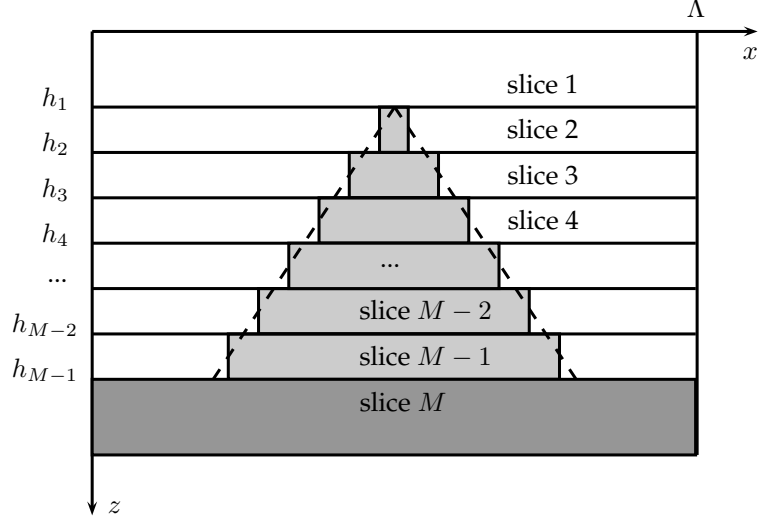
In this chapter we generalize the formulation of the FMM and AFMM-CFF to arbitrary shapes. This implies that, unlike for binary shapes, more than three slices are required in the spatial discretization of the permittivity. We also formulate the spectral discretization in a more elegant way using a Galerkin approach. The cases TE, TM and conical defined in Section 2.4 are discussed.

### 4.1 Standard Fourier modal method

We start the discretization of the time-harmonic Maxwell equations (2.18) by dividing the computational domain vertically into  $M$  slices such that the permittivity may be considered  $z$ -independent in each separate slice. The terminology of "slice" is preferred here to "layer", as the latter is reserved to physical layers in the multilayer stack. As illustrated in Figure 4.1, the upper and lower interface of slice  $l$  are located at  $h_{l-1}$  and  $h_l$  respectively. Since  $z \in \mathbb{R}$ , we take  $h_0 = -\infty$ ,  $h_M = +\infty$ . The permittivities in each slice are given by

$$\epsilon_l(x) = \epsilon(x, z_l), \text{ with } z_l \in [h_{l-1}, h_l]. \quad (4.1)$$

Thus, the profile of the scatterer is approximated by a staircase as in Figure 4.1. The electric and magnetic fields on the computational domain now consist of fields in separate



**Figure 4.1:** Sliced geometry. The dashed line represents the smooth profile being approximated.

slices

$$\mathbf{e}_l(x, y, z) = \mathbf{e}(x, y, z), \quad z \in [h_{l-1}, h_l], \quad (4.2)$$

$$\mathbf{h}_l(x, y, z) = \mathbf{h}(x, y, z), \quad z \in [h_{l-1}, h_l]. \quad (4.3)$$

The time-harmonic Maxwell equations for a slice  $l$  read

$$\frac{\partial}{\partial y} e_{z,l} - \frac{\partial}{\partial z} e_{y,l} = -k_0 h_{x,l}, \quad (4.4a)$$

$$\frac{\partial}{\partial z} e_{x,l} - \frac{\partial}{\partial x} e_{z,l} = -k_0 h_{y,l}, \quad (4.4b)$$

$$\frac{\partial}{\partial x} e_{y,l} - \frac{\partial}{\partial y} e_{x,l} = -k_0 h_{z,l}, \quad (4.4c)$$

$$\frac{1}{\epsilon_l(x)} \frac{\partial}{\partial y} h_{z,l} - \frac{1}{\epsilon_l(x)} \frac{\partial}{\partial z} h_{y,l} = -k_0 e_{x,l}, \quad (4.4d)$$

$$\frac{\partial}{\partial z} h_{x,l} - \frac{\partial}{\partial x} h_{z,l} = -k_0 \epsilon_l(x) e_{y,l}, \quad (4.4e)$$

$$\frac{\partial}{\partial x} h_{y,l} - \frac{\partial}{\partial y} h_{x,l} = -k_0 \epsilon_l(x) e_{z,l}, \quad (4.4f)$$

for  $(x, z) \in [0, \Lambda] \times [h_{i-1}, h_i]$ . Equation (4.4d) has been divided by  $\epsilon_l(x)$  in order to avoid products of functions with concurrent (in the same point) jump discontinuities on the right-hand side. As shown in [41], discretization of the original equation would lead to slower convergence. To discretize in the  $x$  and  $y$  directions we use a Galerkin approach

with "shifted" Fourier harmonics as basis functions and test functions,

$$\phi_n(x, y) = e^{-i(k_{xn}x + k_y y)}, \quad (4.5)$$

where

$$k_{xn} = k_x^{\text{inc}} - n \frac{2\pi}{\Lambda}, \quad k_y = k_y^{\text{inc}}, \quad \text{for } n = -N \dots + N.$$

In each slice  $l$  the fields are expanded as

$$e_{\alpha,l}(x, y, z) = \sum_{n=-N}^N s_{\alpha,l,n}(z) \phi_n(x, y) = (\mathbf{s}_{\alpha,l}(z))^T \cdot \boldsymbol{\phi}(x, y), \quad (4.6a)$$

$$h_{\alpha,l}(x, y, z) = \sum_{n=-N}^N u_{\alpha,l,n}(z) \phi_n(x, y) = (\mathbf{u}_{\alpha,l}(z))^T \cdot \boldsymbol{\phi}(x, y). \quad (4.6b)$$

The  $\alpha$  symbol stands for the  $x$ -,  $y$ -, or  $z$ -component of the field. We apply the Galerkin method with a standard inner product on the interval  $[0, \Lambda]$  to the total field equations (2.18) (see Section B.2 in the Appendix for a detailed derivation). We obtain the discretized equations for the electric and magnetic field,

$$-i\mathbf{K}_y \mathbf{s}_{z,l}(z) - k_0^{-1} \frac{d}{dz} \mathbf{s}_{y,l}(z) = -\mathbf{u}_{x,l}(z), \quad (4.7a)$$

$$k_0^{-1} \frac{d}{dz} \mathbf{s}_{x,l}(z) + i\mathbf{K}_x \mathbf{s}_{z,l}(z) = -\mathbf{u}_{y,l}(z), \quad (4.7b)$$

$$-i\mathbf{K}_x \mathbf{s}_{y,l}(z) + i\mathbf{K}_y \mathbf{s}_{x,l}(z) = -\mathbf{u}_{z,l}(z), \quad (4.7c)$$

$$-i\mathbf{K}_y \mathbf{u}_{z,l}(z) - k_0^{-1} \frac{d}{dz} \mathbf{u}_{y,l}^c(z) = -\mathbf{P}_l^{-1} \mathbf{s}_{x,l}(z), \quad (4.7d)$$

$$k_0^{-1} \frac{d}{dz} \mathbf{u}_{x,l}(z) + i\mathbf{K}_x \mathbf{u}_{z,l}(z) = -\mathbf{E}_l \mathbf{s}_{y,l}(z), \quad (4.7e)$$

$$-i\mathbf{K}_x \mathbf{u}_{y,l}(z) + i\mathbf{K}_y \mathbf{u}_{x,l}(z) = -\mathbf{E}_l \mathbf{s}_{z,l}(z). \quad (4.7f)$$

Introducing the notation for Fourier coefficients  $\hat{\xi}_n$  of a function  $\xi(x)$  on the interval  $x \in [0, \Lambda]$ ,

$$\hat{\xi}_n = \int_0^\Lambda \xi(x) e^{in \frac{2\pi}{\Lambda} x} dx,$$

the matrices in the expressions above are defined as follows,

$$(\mathbf{K}_x)_{mn} = (k_{xn}/k_0) \delta_{mn}, \quad (4.8a)$$

$$(\mathbf{K}_y)_{mn} = (k_y/k_0) \delta_{mn}, \quad (4.8b)$$

$$(\mathbf{E}_l)_{mn} = \hat{\epsilon}_{l,n-m}, \quad (4.8c)$$

$$(\mathbf{P}_l)_{mn} = \hat{p}_{l,n-m}, \quad (4.8d)$$

for  $m, n = -N \dots + N$ . Here  $\delta_{mn}$  is the Kronecker delta and

$$p_l(x) = 1/\epsilon_l(x). \quad (4.9a)$$

The Li rules [41] have been applied in (4.7d). The incident electric and magnetic fields are also represented in terms of the basis functions  $\phi_n$ ,

$$e_\alpha^{\text{inc}}(x, y, z) = \sum_{n=-N}^N s_{\alpha,n}^{\text{inc}}(z) \phi_n(x, y) = (\mathbf{s}_\alpha^{\text{inc}}(z))^T \cdot \boldsymbol{\phi}(x, y), \quad (4.10a)$$

$$h_\alpha^{\text{inc}}(x, y, z) = \sum_{n=-N}^N u_{\alpha,n}^{\text{inc}}(z) \phi_n(x, y) = (\mathbf{u}_\alpha^{\text{inc}}(z))^T \cdot \boldsymbol{\phi}(x, y). \quad (4.10b)$$

It follows from (2.21) and (2.25) that

$$\mathbf{s}_\alpha^{\text{inc}}(z) = a_\alpha^e \mathbf{d}_0 e^{-ik_z^{\text{inc}} z}, \quad (4.11a)$$

$$\mathbf{u}_\alpha^{\text{inc}}(z) = a_\alpha^h \mathbf{d}_0 e^{-ik_z^{\text{inc}} z}. \quad (4.11b)$$

### 4.1.1 TE-polarization

In the case of planar incidence and TE-polarization we have  $e_x = e_z = h_y = 0$ , which implies that  $\mathbf{s}_x = \mathbf{s}_z = \mathbf{u}_y = \mathbf{0}$ . The discretized Maxwell equations (4.7) reduce to

$$-k_0^{-1} \frac{d}{dz} \mathbf{s}_{y,l}(z) = -\mathbf{u}_{x,l}(z), \quad (4.12a)$$

$$-i\mathbf{K}_x \mathbf{s}_{y,l}(z) = -\mathbf{u}_{z,l}(z), \quad (4.12b)$$

$$k_0^{-1} \frac{d}{dz} \mathbf{u}_{x,l}(z) + i\mathbf{K}_x \mathbf{u}_{z,l}(z) = -\mathbf{E}_l \mathbf{s}_{y,l}(z). \quad (4.12c)$$

After elimination of  $\mathbf{u}_x$  and  $\mathbf{u}_z$  we are left with an equation for  $\mathbf{s}_y$

$$\frac{d^2}{dz^2} \mathbf{s}_{y,l}(z) = k_0^2 \mathbf{A}_l \mathbf{s}_{y,l}(z), \quad (4.13)$$

where  $\mathbf{A}_l = \mathbf{K}_x^2 - \mathbf{E}_l$ . Equation (4.13) is a system of homogeneous second-order ordinary differential equations. Its general solution is given by

$$\mathbf{s}_{y,l}(z) = \mathbf{W}_l (e^{-k_0 \mathbf{Q}_l (z-h_{l-1})} \mathbf{c}_l^+ + e^{k_0 \mathbf{Q}_l (z-h_l)} \mathbf{c}_l^-) = \mathbf{s}_{y,l}^+(z) + \mathbf{s}_{y,l}^-(z), \quad (4.14)$$

where  $\mathbf{W}_l$  is the matrix of eigenvectors of  $\mathbf{A}_l$ , and  $\mathbf{Q}_l$  is a diagonal matrix with square roots of the corresponding eigenvalues on its diagonal. The quantities with  $^+$  superscript correspond to waves traveling in the positive  $z$ -direction (downward), and quantities with  $^-$  superscript correspond to waves traveling in the negative  $z$ -direction (upward).

At the interface, continuity of the tangential components of the fields is required,

$$\mathbf{s}_{y,l}(h_l) = \mathbf{s}_{y,l+1}(h_l), \quad (4.15a)$$

$$\mathbf{u}_{x,l}(h_l) = \mathbf{u}_{x,l+1}(h_l). \quad (4.15b)$$

These conditions hold for the contrast field as a result of the continuity of tangential components of the total and background fields. Using (4.12a) yields

$$\mathbf{s}_{y,l}(h_l) = \mathbf{s}_{y,l+1}(h_l), \quad (4.16a)$$

$$k_0^{-1} \frac{d}{dz} \mathbf{s}_{y,l}(h_l) = k_0^{-1} \frac{d}{dz} \mathbf{s}_{y,l+1}(h_l). \quad (4.16b)$$

We define  $\mathbf{X}_l = e^{-k_0 \mathbf{Q}_l (h_l - h_{l-1})}$  and  $\mathbf{V}_l = -\mathbf{W}_l \mathbf{Q}_l$ . Then, from (4.16) and (4.14) we have for each slice

$$\begin{bmatrix} \mathbf{W}_l \mathbf{X}_l & \mathbf{W}_l \\ \mathbf{V}_l \mathbf{X}_l & -\mathbf{V}_l \end{bmatrix} \begin{bmatrix} \mathbf{c}_l^+ \\ \mathbf{c}_l^- \end{bmatrix} = \begin{bmatrix} \mathbf{W}_{l+1} & \mathbf{W}_{l+1} \mathbf{X}_{l+1} \\ \mathbf{V}_{l+1} & -\mathbf{V}_{l+1} \mathbf{X}_{l+1} \end{bmatrix} \begin{bmatrix} \mathbf{c}_{l+1}^+ \\ \mathbf{c}_{l+1}^- \end{bmatrix}. \quad (4.17)$$

The radiation condition needs to be applied in the top and bottom slices. Because the only incoming field in slice 1 is the incident field (illumination), we impose a restriction on the term representing the downward traveling waves in this slice (see the general solution (4.14))

$$\mathbf{s}_{y,1}^+ = \mathbf{s}_y^{\text{inc}}. \quad (4.18)$$

From (2.29) we determine  $\mathbf{s}_y^{\text{inc}} = \mathbf{d}_0 e^{-ik_z^{\text{inc}} z}$ . In the linear system (4.17) (which results from the interface conditions (4.15)) we have

$$\begin{bmatrix} \mathbf{W}_1 \mathbf{X}_1 \\ \mathbf{V}_1 \mathbf{X}_1 \end{bmatrix} \mathbf{c}_1^+ = \begin{bmatrix} \mathbf{d}_0 \\ -ik_0^{-1} k_z^{\text{inc}} \mathbf{d}_0 \end{bmatrix} e^{-ik_z^{\text{inc}} h_1}. \quad (4.19)$$

In slice  $M$  there is no incoming field:

$$\mathbf{s}_{y,M}^- = \mathbf{0}. \quad (4.20)$$

Thus in the linear system (4.17) we have

$$\begin{bmatrix} \mathbf{W}_M \\ \mathbf{V}_M \end{bmatrix} \mathbf{c}_M^- = \mathbf{0}. \quad (4.21)$$

### 4.1.2 TM-polarization

In the case of planar incidence and TM-polarization we have  $e_y = h_x = h_z = 0$ , which implies that  $\mathbf{s}_y = \mathbf{u}_x = \mathbf{u}_z = \mathbf{0}$ . The Maxwell equations (4.7) reduce to

$$k_0^{-1} \frac{d}{dz} \mathbf{s}_{x,l}(z) + i\mathbf{K}_x \mathbf{s}_{z,l}(z) = -\mathbf{u}_{y,l}(z), \quad (4.22a)$$

$$-k_0^{-1} \mathbf{P}_l \frac{d}{dz} \mathbf{u}_{y,l}(z) = -\mathbf{s}_{x,l}(z), \quad (4.22b)$$

$$-i\mathbf{K}_x \mathbf{u}_{y,l}(z) = -\mathbf{E}_l \mathbf{s}_{z,l}(z). \quad (4.22c)$$

After elimination of  $\mathbf{s}_x$  and  $\mathbf{s}_z$  we are left with an equation for  $\mathbf{u}_y$

$$\frac{d^2}{dz^2} \mathbf{u}_{y,l} = \mathbf{P}_l^{-1} \mathbf{B}_l \mathbf{u}_{y,l}, \quad (4.23)$$

where  $\mathbf{B}_l = (\mathbf{K}_x \mathbf{E}_l^{-1} \mathbf{K}_x - \mathbf{I})$ . Similarly to the TE-polarization case, the general solution is given by

$$\mathbf{u}_{y,l}(z) = \mathbf{W}_l (e^{-k_0 \mathbf{Q}_l (z-h_{l-1})} \mathbf{c}_l^+ + e^{k_0 \mathbf{Q}_l (z-h_l)} \mathbf{c}_l^-), \quad (4.24)$$

where  $\mathbf{W}_l$  and  $\mathbf{Q}_l$  are respectively the matrix of eigenvectors and diagonal matrix with square roots of eigenvalues of  $\mathbf{P}_l^{-1} \mathbf{B}_l$  on the diagonal. At the interface, continuity of the tangential components of the fields is required,

$$\mathbf{u}_{y,l}(h_l) = \mathbf{u}_{y,l+1}(h_l), \quad (4.25a)$$

$$\mathbf{s}_{x,l}(h_l) = \mathbf{s}_{x,l+1}(h_l). \quad (4.25b)$$

Using (4.22b) yields

$$\mathbf{u}_{y,l}(h_l) = \mathbf{u}_{y,l+1}(h_l), \quad (4.26a)$$

$$k_0^{-1} \mathbf{P}_l \frac{d}{dz} \mathbf{u}_{y,l}(h_l) = k_0^{-1} \mathbf{P}_{l+1} \frac{d}{dz} \mathbf{u}_{y,l+1}(h_l).$$

We define  $\mathbf{V}_l = -\mathbf{P}_l \mathbf{W}_l \mathbf{Q}_l$  (keeping the old definition of  $\mathbf{X}_l$ ). Then, from (4.26) and (4.24) we have for each slice

$$\begin{bmatrix} \mathbf{W}_l \mathbf{X}_l & \mathbf{W}_l \\ \mathbf{V}_l \mathbf{X}_l & -\mathbf{V}_l \end{bmatrix} \begin{bmatrix} \mathbf{c}_l^+ \\ \mathbf{c}_l^- \end{bmatrix} = \begin{bmatrix} \mathbf{W}_{l+1} & \mathbf{W}_{l+1} \mathbf{X}_{l+1} \\ \mathbf{V}_{l+1} & -\mathbf{V}_{l+1} \mathbf{X}_{l+1} \end{bmatrix} \begin{bmatrix} \mathbf{c}_{l+1}^+ \\ \mathbf{c}_{l+1}^- \end{bmatrix}. \quad (4.27)$$

We apply the radiation conditions in the top and bottom slices in a similar fashion to the TE-case. In slice 1, we require that

$$\mathbf{u}_{y,1}^+ = \mathbf{u}_y^{\text{inc}}. \quad (4.28)$$

From (2.32) we determine  $\mathbf{u}_y^{\text{inc}} = in_1 \mathbf{d}_0 e^{-ik_z^{\text{inc}} z}$ . In the linear system (4.27) we have

$$\begin{bmatrix} \mathbf{W}_1 \mathbf{X}_1 \\ \mathbf{V}_1 \mathbf{X}_1 \end{bmatrix} \mathbf{c}_1^+ = \begin{bmatrix} in_1 \mathbf{d}_0 \\ k_0^{-1} k_z^{\text{inc}} \mathbf{d}_0 \end{bmatrix} e^{-ik_z^{\text{inc}} h_1}. \quad (4.29)$$

In slice  $M$  there is no incoming field:

$$\mathbf{u}_{y,M}^- = \mathbf{0}. \quad (4.30)$$

Thus in the linear system (4.17) we have

$$\begin{bmatrix} \mathbf{W}_M \mathbf{X}_M \\ -\mathbf{V}_M \mathbf{X}_M \end{bmatrix} \mathbf{c}_M^- = \mathbf{0}. \quad (4.31)$$

### 4.1.3 Conical incidence

The case of conical incidence is the most general as it allows all the components of the fields to be non-zero. The discretized Maxwell equations (4.7) can be reduced to two second-order ordinary differential equations for  $\mathbf{s}_{x,l}$  and  $\mathbf{u}_{x,l}$ . We briefly describe the derivation of the equation for the latter.

Using (4.7f) and (4.7e) in (4.7a) yields

$$\mathbf{K}_y \mathbf{E}_l^{-1} \mathbf{K}_x \mathbf{u}_{y,l} - \mathbf{K}_y^2 \mathbf{E}_l^{-1} \mathbf{u}_{x,l} + k_0^{-2} \mathbf{E}_l^{-1} \frac{d^2}{dz^2} \mathbf{u}_{x,l} + ik_0^{-1} \mathbf{E}_l^{-1} \mathbf{K}_x \frac{d}{dz} \mathbf{u}_{z,l} = -\mathbf{u}_{x,l}. \quad (4.32)$$

The first and last terms in the left-hand side are expressed from (4.7b) and (4.7c) respectively

$$\begin{aligned} & \mathbf{E}_l^{-1} \mathbf{K}_x (\mathbf{K}_y \mathbf{u}_{y,l} + ik_0^{-1} \frac{d}{dz} \mathbf{u}_{z,l}) \\ &= \mathbf{E}_l^{-1} \mathbf{K}_x (-k_0^{-1} \mathbf{K}_y \frac{d}{dz} \mathbf{s}_{x,l} - i \mathbf{K}_y \mathbf{K}_x \mathbf{s}_{x,l} - k_0^{-1} \mathbf{K}_x \frac{d}{dz} \mathbf{s}_{y,l} + k_0^{-1} \mathbf{K}_y \frac{d}{dz} \mathbf{s}_{x,l}) \\ &= \mathbf{E}_l^{-1} \mathbf{K}_x^2 (-i \mathbf{K}_y \mathbf{s}_{x,l} - k_0^{-1} \frac{d}{dz} \mathbf{s}_{y,l}) \\ &= -\mathbf{E}_l^{-1} \mathbf{K}_x^2 \mathbf{u}_{x,l}, \end{aligned} \quad (4.33)$$

where in the last step Equation (4.7a) was used. Substitution of (4.33) in (4.32) gives

$$-\mathbf{E}_l^{-1} \mathbf{K}_x^2 \mathbf{u}_{x,l} - \mathbf{K}_y^2 \mathbf{E}_l^{-1} \mathbf{u}_{x,l} + k_0^{-2} \mathbf{E}_l^{-1} \frac{d^2}{dz^2} \mathbf{u}_{x,l} = -\mathbf{u}_{x,l}. \quad (4.34)$$



Finally we arrive at a homogeneous second-order ODE:

$$\frac{d^2}{dz^2} \mathbf{u}_{x,l} = (\mathbf{K}_x^2 + \mathbf{K}_y^2 - \mathbf{E}_l) \mathbf{u}_{x,l}. \quad (4.35)$$

The second-order equation for  $\mathbf{s}_{x,l}$  is derived in a similar manner, so that we obtain

$$\frac{d^2}{dz^2} \mathbf{s}_{x,l}(z) = k_0^2 \mathbf{C}_l \mathbf{s}_{x,l}(z), \quad (4.36a)$$

$$\frac{d^2}{dz^2} \mathbf{u}_{x,l}(z) = k_0^2 \mathbf{D}_l \mathbf{u}_{x,l}(z), \quad (4.36b)$$

where

$$\mathbf{C}_l = \mathbf{K}_y^2 + (\mathbf{K}_x \mathbf{E}_l^{-1} \mathbf{K}_x - \mathbf{I}) \mathbf{P}_l^{-1} = \mathbf{K}_y^2 + \mathbf{B}_l \mathbf{P}_l^{-1}, \quad (4.37)$$

$$\mathbf{D}_l = \mathbf{K}_y^2 + \mathbf{K}_x^2 - \mathbf{E}_l = \mathbf{K}_y^2 + \mathbf{A}_l. \quad (4.38)$$

The general solution of (4.36) is given by

$$\mathbf{s}_{x,l}(z) = \mathbf{W}_{s,l} (e^{-k_0 \mathbf{Q}_{s,l} (z-h_{l-1})} \mathbf{c}_{s,l}^+ + e^{k_0 \mathbf{Q}_{s,l} (z-h_l)} \mathbf{c}_{s,l}^-), \quad (4.39a)$$

$$\mathbf{u}_{x,l}(z) = \mathbf{W}_{u,l} (e^{-k_0 \mathbf{Q}_{u,l} (z-h_{l-1})} \mathbf{c}_{u,l}^+ + e^{k_0 \mathbf{Q}_{u,l} (z-h_l)} \mathbf{c}_{u,l}^-), \quad (4.39b)$$

where the pairs  $\mathbf{W}_{s,l}$ ,  $\mathbf{Q}_{s,l}$  and  $\mathbf{W}_{u,l}$ ,  $\mathbf{Q}_{u,l}$  contain the matrix of eigenvectors and the diagonal matrix with square roots of eigenvalues of, respectively,  $\mathbf{C}_l$  and  $\mathbf{D}_l$  on the diagonal.

At the interface, continuity of the tangential components of the fields is required

$$\mathbf{s}_{x,l}(h_l) = \mathbf{s}_{x,l+1}(h_l), \quad (4.40a)$$

$$\mathbf{s}_{y,l}(h_l) = \mathbf{s}_{y,l+1}(h_l), \quad (4.40b)$$

$$\mathbf{u}_{x,l}(h_l) = \mathbf{u}_{x,l+1}(h_l), \quad (4.40c)$$

$$\mathbf{u}_{y,l}(h_l) = \mathbf{u}_{y,l+1}(h_l). \quad (4.40d)$$

Using (4.7c) in (4.7e) and (4.7f) in (4.7b), the  $y$ -components of the fields are expressed in terms of  $x$ -components,

$$(\mathbf{K}_x^2 - \mathbf{E}_l) \mathbf{s}_{y,l} = \mathbf{K}_x \mathbf{K}_y \mathbf{s}_{x,l} + k_0^{-1} \frac{d}{dz} \mathbf{u}_{x,l}, \quad (4.41a)$$

$$(\mathbf{K}_x \mathbf{E}_l^{-1} \mathbf{K}_x - \mathbf{I}) \mathbf{u}_{y,l} = \mathbf{K}_x \mathbf{E}_l^{-1} \mathbf{K}_y \mathbf{u}_{x,l} + k_0^{-1} \frac{d}{dz} \mathbf{s}_{x,l}. \quad (4.41b)$$

Finally we obtain

$$\mathbf{s}_{y,l} = \mathbf{A}_l^{-1} \left( \mathbf{K}_x \mathbf{K}_y \mathbf{s}_{x,l} + k_0^{-1} \frac{d}{dz} \mathbf{u}_{x,l} \right), \quad (4.42a)$$

$$\mathbf{u}_{y,l} = \mathbf{B}_l^{-1} \left( \mathbf{K}_x \mathbf{E}^{-1} \mathbf{K}_y \mathbf{u}_{x,l} + k_0^{-1} \frac{d}{dz} \mathbf{s}_{x,l} \right). \quad (4.42b)$$

We define

$$\mathbf{W}_l = \begin{bmatrix} \mathbf{0} & \mathbf{W}_{s,l} \\ \mathbf{A}_l^{-1} \mathbf{W}_{u,l} \mathbf{Q}_{u,l} & \mathbf{A}_l^{-1} \mathbf{K}_x \mathbf{K}_y \mathbf{W}_{s,l} \end{bmatrix}, \quad (4.43)$$

$$\mathbf{V}_l = - \begin{bmatrix} \mathbf{W}_{u,l} & \mathbf{0} \\ \mathbf{B}_l^{-1} \mathbf{K}_x \mathbf{E}_l^{-1} \mathbf{K}_y \mathbf{W}_{u,l} & \mathbf{B}_l^{-1} \mathbf{W}_{s,l} \mathbf{Q}_{s,l} \end{bmatrix}, \quad (4.44)$$

and

$$\mathbf{c}_l^+ = \begin{bmatrix} -\mathbf{c}_{u,l}^+ \\ \mathbf{c}_{s,l}^+ \end{bmatrix}, \quad \mathbf{c}_l^- = \begin{bmatrix} \mathbf{c}_{u,l}^- \\ \mathbf{c}_{s,l}^- \end{bmatrix}. \quad (4.45)$$

Then, from (4.40) and (4.39) we have for each slice

$$\begin{bmatrix} \mathbf{W}_l \mathbf{X}_l & \mathbf{W}_l \\ \mathbf{V}_l \mathbf{X}_l & -\mathbf{V}_l \end{bmatrix} \begin{bmatrix} \mathbf{c}_l^+ \\ \mathbf{c}_l^- \end{bmatrix} = \begin{bmatrix} \mathbf{W}_{l+1} & \mathbf{W}_{l+1} \mathbf{X}_{l+1} \\ \mathbf{V}_{l+1} & -\mathbf{V}_{l+1} \mathbf{X}_{l+1} \end{bmatrix} \begin{bmatrix} \mathbf{c}_{l+1}^+ \\ \mathbf{c}_{l+1}^- \end{bmatrix}. \quad (4.46)$$

In order to apply the radiation condition in slice 1, we require that

$$\begin{bmatrix} \mathbf{s}_{x,1}^+(z) \\ \mathbf{s}_{y,1}^+(z) \\ \mathbf{u}_{x,1}^+(z) \\ \mathbf{u}_{y,1}^+(z) \end{bmatrix} = \begin{bmatrix} \mathbf{s}_x^{\text{inc}}(z) \\ \mathbf{s}_y^{\text{inc}}(z) \\ \mathbf{u}_x^{\text{inc}}(z) \\ \mathbf{u}_y^{\text{inc}}(z) \end{bmatrix} \stackrel{(4.11)}{=} \begin{bmatrix} a_x^e \mathbf{d}_0 \\ a_y^e \mathbf{d}_0 \\ a_x^h \mathbf{d}_0 \\ a_y^h \mathbf{d}_0 \end{bmatrix} e^{-ik_z^{\text{inc}} z}. \quad (4.47)$$

In the linear system (4.46) we have

$$\begin{bmatrix} \mathbf{W}_1 \mathbf{X}_1 \\ \mathbf{V}_1 \mathbf{X}_1 \end{bmatrix} \mathbf{c}_1^+ = \begin{bmatrix} a_x^e \mathbf{d}_0 \\ a_y^e \mathbf{d}_0 \\ a_x^h \mathbf{d}_0 \\ a_y^h \mathbf{d}_0 \end{bmatrix} e^{-ik_z^{\text{inc}} h_1}. \quad (4.48)$$

In slice  $M$  there is no incoming field:

$$\begin{bmatrix} \mathbf{s}_{x,M}^- \\ \mathbf{s}_{y,M}^- \\ \mathbf{u}_{x,M}^- \\ \mathbf{u}_{y,M}^- \end{bmatrix} = \mathbf{0}. \quad (4.49)$$

Thus in the linear system (4.46) we have

$$\begin{bmatrix} \mathbf{W}_M \mathbf{X}_M \\ -\mathbf{V}_M \mathbf{X}_M \end{bmatrix} \mathbf{c}_M^- = \mathbf{0}. \quad (4.50)$$

## 4.2 Aperiodic Fourier modal method

In the AFMM-CFF we use PMLs [5] in order to impose the radiation condition at the lateral boundaries. The radiation condition however imposes a restriction on the problems which can be solved: no incident field is allowed. Therefore, the Maxwell equations need to be reformulated such that the incident field is replaced by a virtual source. To this end, an associated background problem is defined,

$$\nabla \times \mathbf{e}^b(\mathbf{x}) = -k_0 \mathbf{h}^b(\mathbf{x}), \quad (4.51a)$$

$$\nabla \times \mathbf{h}^b(\mathbf{x}) = -k_0 \epsilon^b(x, z) \mathbf{e}^b(\mathbf{x}), \quad (4.51b)$$

with

$$\mathbf{e}^{b,\text{inc}}(\mathbf{x}) = \mathbf{a} e^{-i\mathbf{k}^{\text{inc}} \cdot \mathbf{x}}. \quad (4.51c)$$

This problem is chosen such that it admits an analytical solution and the function  $\epsilon - \epsilon^b$  has compact support. As explained in Chapter 3 (see also [59, 60, 61]), the compact support condition is required in order to avoid non-zero source terms in the PML. Typically  $\epsilon^b$  represents the permittivity of the background multilayer which supports the scatterer. Subtraction of (4.51) from (2.17) yields the *contrast-field formulation*

$$\nabla \times \mathbf{e}^c(\mathbf{x}) = -k_0 \mathbf{h}^c(\mathbf{x}), \quad (4.52a)$$

$$\nabla \times \mathbf{h}^c(\mathbf{x}) = -k_0 \epsilon(x, z) \mathbf{e}^c(\mathbf{x}) - k_0 (\epsilon(x, z) - \epsilon^b(x, z)) \mathbf{e}^b(\mathbf{x}), \quad (4.52b)$$

with

$$\mathbf{e}^{c,\text{inc}}(\mathbf{x}) = \mathbf{0}. \quad (4.52c)$$

The PMLs can be viewed as an analytical continuation of the solution into the complex plane [11, 13]. For PMLs placed in the  $x$ -direction, this implies a change of the  $x$ -derivative in the differential equations (4.52):

$$\frac{\partial}{\partial x} \rightarrow \frac{1}{f'(x)} \frac{\partial}{\partial x}, \quad \text{with } f(x) = x + i\beta(x). \quad (4.53)$$

The function  $\beta$  is continuous and non-zero only in the PMLs, which are placed in the stripes  $x \in [0, x_l]$  and  $x \in [x_r, \Lambda]$ . In Chapter 3 the tilde ( $\tilde{\phantom{x}}$ ) notation was used for

quantities related to the aperiodic problem to achieve clear distinction from the periodic problem. For convenience this notation is abandoned in this and following chapters. We mention that although PMLs are originally designed for hyperbolic equations, they have also been recently successfully applied to the heat equation [37]. Lalanne and co-workers [36] were the first to use PMLs with Fourier modal methods. They chose trigonometric stretching functions whose Fourier coefficients can be computed analytically. Many other forms for  $f$  (and implicitly for  $\beta$ ) have been suggested [6, 12, 57]. Typically, a polynomial or geometric variation of  $f$  in the PML is used. We adopt the first form. An example of such a function was shown in Figure 3.2.

In this chapter we introduce a slightly different PML than the one defined by (3.47) in Chapter 3. It is given by the coordinate transformation

$$f(x) = \begin{cases} x + ik_0^p(\beta_0|x - x_l|)^{p+1}, & 0 \leq x \leq x_l, \\ x, & x_l < x < x_r, \\ x - ik_0^p(\beta_0|x - x_r|)^{p+1}, & x_r \leq x \leq \Lambda, \end{cases} \quad (4.54)$$

where  $x_l$  is the right end of the left PML,  $x_r$  is the left end of the right PML, and  $\beta_0$  is the damping strength. To the authors' knowledge, no study exists on the choice of  $p$  for PMLs in the aperiodic Fourier modal methods. For instance, for FDTD methods nearly optimal results have been obtained for  $p \in [3, 4]$  [6, 90]. In our computations we set  $p = 1$ .

We look at the amplitude of a plane wave with unit amplitude after passing through the PML. Let  $x \in [x_r, \Lambda)$  be a point in the PML and  $\Delta x = x - x_r$ . We have

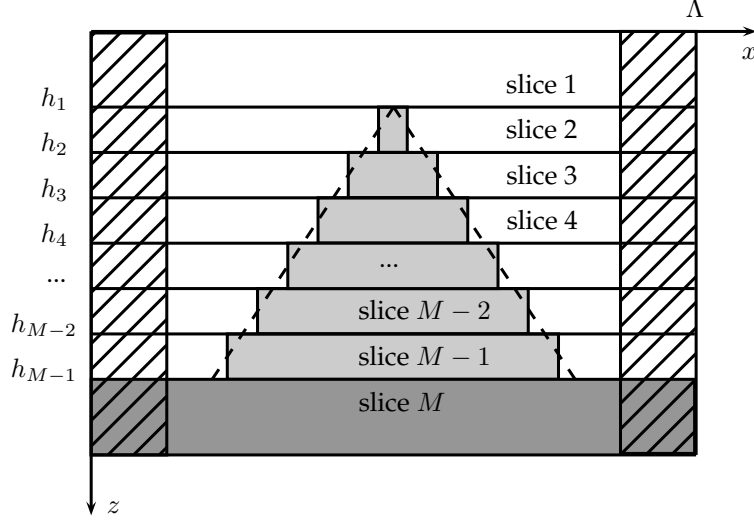
$$|e^{-ik_0f(x)}| = |e^{-ik_0x}| |e^{-(k_0\beta_0\Delta x)^{p+1}}| = e^{-(k_0\beta_0\Delta x)^{p+1}}.$$

Now, we determine the distance over which the amplitude decays from 1 to  $e^{-1}$ ,

$$-(k_0\beta_0\Delta x)^{p+1} = -1 \Rightarrow k_0\Delta x = \frac{1}{\beta_0}.$$

This relation provides an intuitive meaning for  $\beta_0$  - it determines the *inverse decay length*. Since  $k_0 = 2\pi/\lambda_0$ , the length is scaled by the wavelength of the incident field.

Similarly to the discretization in the standard FMM, we divide the domain into  $M$  slices



**Figure 4.2:** Sliced geometry. The dashed line represents the smooth profile being approximated. The hatched areas indicate PMLs.

(see Figure 4.2). Then Maxwell equations (4.52) for the contrast field in slice  $l$  read

$$\frac{\partial}{\partial y} e_{z,l}^c - \frac{\partial}{\partial z} e_{y,l}^c = -k_0 h_{x,l}^c, \quad (4.55a)$$

$$\frac{\partial}{\partial z} e_{x,l}^c - \frac{1}{f'(x)} \frac{\partial}{\partial x} e_{z,l}^c = -k_0 h_{y,l}^c, \quad (4.55b)$$

$$\frac{1}{f'(x)} \frac{\partial}{\partial x} e_{y,l}^c - \frac{\partial}{\partial y} e_{x,l}^c = -k_0 h_{z,l}^c, \quad (4.55c)$$

$$\frac{1}{\epsilon_l(x)} \frac{\partial}{\partial y} h_{z,l}^c - \frac{1}{\epsilon_l(x)} \frac{\partial}{\partial z} h_{y,l}^c = -k_0 e_{x,l}^c - k_0 \left(1 - \frac{1}{\epsilon_l(x)} \epsilon_l^b\right) e_{x,l}^b, \quad (4.55d)$$

$$\frac{\partial}{\partial z} h_{x,l}^c - \frac{1}{f'(x)} \frac{\partial}{\partial x} h_{z,l}^c = -k_0 \epsilon_l(x) e_{y,l}^c - k_0 (\epsilon_l(x) - \epsilon_l^b) e_{y,l}^b, \quad (4.55e)$$

$$\frac{1}{f'(x)} \frac{\partial}{\partial x} h_{y,l}^c - \frac{\partial}{\partial y} h_{x,l}^c = -k_0 \epsilon_l(x) e_{z,l}^c - k_0 (\epsilon_l(x) - \epsilon_l^b) e_{z,l}^b, \quad (4.55f)$$

for  $(x, z) \in [0, \Lambda] \times [h_{l-1}, h_l]$ . As it was done in Equation (4.4d), Equation (4.55d) has been divided by  $\epsilon_l(x)$  in order to avoid products of functions with concurrent (in the same point) jump discontinuities on the right-hand side.

In the  $x$ -direction we use a Galerkin approach with "shifted" Fourier harmonics (4.5) as basis functions and test functions. In each slice  $l$  the contrast (electric and magnetic)

fields are expanded as

$$e_{\alpha,l}^{c,N}(x, y, z) = \sum_{n=-N}^N s_{\alpha,l,n}^c(z) \phi_n(x, y) = (\mathbf{s}_{\alpha,l}^c(z))^T \cdot \boldsymbol{\phi}(x, y), \quad (4.56a)$$

$$h_{\alpha,l}^{c,N}(x, y, z) = \sum_{n=-N}^N u_{\alpha,l,n}^c(z) \phi_n(x, y) = (\mathbf{u}_{\alpha,l}^c(z))^T \cdot \boldsymbol{\phi}(x, y). \quad (4.56b)$$

Again, the  $\alpha$  symbol stands for the  $x$ -,  $y$ -, or  $z$ -component of the field. The background fields, which are known in advance, have to be represented in the same basis as the contrast field, i.e.

$$e_{\alpha,l}^{b,N}(x, y, z) = \sum_{n=-N}^N s_{\alpha,l,n}^b(z) \phi_n(x, y) = (\mathbf{s}_{\alpha,l}^b(z))^T \cdot \boldsymbol{\phi}(x, y), \quad (4.57a)$$

$$h_{\alpha,l}^{b,N}(x, y, z) = \sum_{n=-N}^N u_{\alpha,l,n}^b(z) \phi_n(x, y) = (\mathbf{u}_{\alpha,l}^b(z))^T \cdot \boldsymbol{\phi}(x, y). \quad (4.57b)$$

The background field is determined in advance by solving the Fresnel reflection-transmission problem for a multilayer.

$$\mathbf{s}_{\alpha,l}^b(z) = \mathbf{d}_0 s_{\alpha,l}^b(z) = \mathbf{d}_0 \left( a_{\alpha,l}^e e^{-k_0 q_l (z-h_{l-1})} + r_{\alpha,l}^e e^{k_0 q_l (z-h_l)} \right), \quad (4.58a)$$

$$\mathbf{u}_{\alpha,l}^b(z) = \mathbf{d}_0 u_{\alpha,l}^b(z) = \mathbf{d}_0 \left( a_{\alpha,l}^h e^{-k_0 q_l (z-h_{l-1})} + r_{\alpha,l}^h e^{k_0 q_l (z-h_l)} \right), \quad (4.58b)$$

where  $\mathbf{d}_0 \in \mathbb{R}^{2N+1}$  is an all-zero vector except for entry  $N+1$  and  $q_l$  is defined as

$$q_l = i \sqrt{\epsilon_l^b - \left( \frac{k_x^{\text{inc}}}{k_0} \right)^2 - \left( \frac{k_y^{\text{inc}}}{k_0} \right)^2}. \quad (4.59)$$

The coefficients  $a_{\alpha,l}^e$ ,  $r_{\alpha,l}^e$ , and  $a_{\alpha,l}^h$ ,  $r_{\alpha,l}^h$  in (4.58) are the amplitudes of the downward and upward traveling waves corresponding to the electric and magnetic background field.

We apply the Galerkin method with a standard inner product on the interval  $x \in [0, \Lambda)$

to the contrast field equations (4.55)

$$-i\mathbf{K}_y \mathbf{s}_{z,l}^c(z) - k_0^{-1} \frac{d}{dz} \mathbf{s}_{y,l}^c(z) = -\mathbf{u}_{x,l}^c(z), \quad (4.60a)$$

$$k_0^{-1} \frac{d}{dz} \mathbf{s}_{x,l}^c(z) + i\mathbf{F}\mathbf{K}_x \mathbf{s}_{z,l}^c(z) = -\mathbf{u}_{y,l}^c(z), \quad (4.60b)$$

$$-i\mathbf{F}\mathbf{K}_x \mathbf{s}_{y,l}^c(z) + i\mathbf{K}_y \mathbf{s}_{x,l}^c(z) = -\mathbf{u}_{z,l}^c(z), \quad (4.60c)$$

$$-i\mathbf{K}_y \mathbf{u}_{z,l}^c(z) - k_0^{-1} \frac{d}{dz} \mathbf{u}_{y,l}^c(z) = -\mathbf{P}_l^{-1} \mathbf{s}_{x,l}^c(z) - (\mathbf{P}_l^{-1} - (\mathbf{P}_l^b)^{-1}) \mathbf{s}_{x,l}^b(z), \quad (4.60d)$$

$$k_0^{-1} \frac{d}{dz} \mathbf{u}_{x,l}^c(z) + i\mathbf{F}\mathbf{K}_x \mathbf{u}_{z,l}^c(z) = -\mathbf{E}_l \mathbf{s}_{y,l}^c(z) - (\mathbf{E}_l - \mathbf{E}_l^b) \mathbf{s}_{y,l}^b(z), \quad (4.60e)$$

$$-i\mathbf{F}\mathbf{K}_x \mathbf{u}_{y,l}^c(z) + i\mathbf{K}_y \mathbf{u}_{x,l}^c(z) = -\mathbf{E}_l \mathbf{s}_{z,l}^c(z) - (\mathbf{E}_l - \mathbf{E}_l^b) \mathbf{s}_{z,l}^b(z). \quad (4.60f)$$

The matrices in the expressions above are defined as follows,

$$(\mathbf{K}_x)_{mn} = (k_{xn}/k_0) \delta_{mn}, \quad (4.61a)$$

$$(\mathbf{K}_y)_{mn} = (k_{yn}/k_0) \delta_{mn}, \quad (4.61b)$$

$$(\mathbf{E}_l)_{mn} = \hat{\epsilon}_{l,n-m}, \quad (4.61c)$$

$$(\mathbf{P}_l)_{mn} = \hat{p}_{l,n-m}, \quad (4.61d)$$

$$(\mathbf{E}_l^b)_{mn} = \hat{\epsilon}_{l,n-m}^b, \quad (4.61e)$$

$$(\mathbf{P}_l^b)_{mn} = \hat{p}_{l,n-m}^b, \quad (4.61f)$$

$$(\mathbf{F})_{mn} = \hat{\gamma}_{n-m}, \quad (4.61g)$$

for  $m, n = -N \dots + N$ . Here  $\delta_{mn}$  is the Kronecker delta and

$$p_l(x) = 1/\epsilon_l(x), \quad (4.62a)$$

$$p_l^b = 1/\epsilon_l^b, \quad (4.62b)$$

$$\gamma(x) = 1/f'(x). \quad (4.62c)$$

The function  $f(x)$  is the complex coordinate transformation implementing the PML. Since  $\epsilon_l^b$  and  $p_l^b$  are  $x$ -independent, the matrices  $\mathbf{E}_l^b = \epsilon_l^b \mathbf{I}$  and  $\mathbf{P}_l^b = (\epsilon_l^b)^{-1} \mathbf{I}$  are diagonal.

For future reference we also list the discretized Maxwell equations for the background

field:

$$-i\mathbf{K}_y \mathbf{s}_{z,l}^b(z) - k_0^{-1} \frac{d}{dz} \mathbf{s}_{y,l}^b(z) = -\mathbf{u}_{x,l}^b(z), \quad (4.63a)$$

$$k_0^{-1} \frac{d}{dz} \mathbf{s}_{x,l}^b(z) + i\mathbf{F}\mathbf{K}_x \mathbf{s}_{z,l}^b(z) = -\mathbf{u}_{y,l}^b(z), \quad (4.63b)$$

$$-i\mathbf{F}\mathbf{K}_x \mathbf{s}_{y,l}^b(z) + i\mathbf{K}_y \mathbf{s}_{x,l}^b(z) = -\mathbf{u}_{z,l}^b(z), \quad (4.63c)$$

$$-i\mathbf{K}_y \mathbf{u}_{z,l}^b(z) - k_0^{-1} \frac{d}{dz} \mathbf{u}_{y,l}^b(z) = -(\mathbf{P}_l^b)^{-1} \mathbf{s}_{x,l}^b(z), \quad (4.63d)$$

$$k_0^{-1} \frac{d}{dz} \mathbf{u}_{x,l}^b(z) + i\mathbf{F}\mathbf{K}_x \mathbf{u}_{z,l}^b(z) = -\mathbf{E}_l^b \mathbf{s}_{y,l}^b(z), \quad (4.63e)$$

$$-i\mathbf{F}\mathbf{K}_x \mathbf{u}_{y,l}^b(z) + i\mathbf{K}_y \mathbf{u}_{x,l}^b(z) = -\mathbf{E}_l^b \mathbf{s}_{z,l}^b(z). \quad (4.63f)$$

### 4.2.1 TE-polarization

In the case of planar incidence and TE-polarization we have  $\mathbf{s}_x^c = \mathbf{s}_z^c = \mathbf{u}_y^c = \mathbf{0}$  and  $\mathbf{s}_x^b = \mathbf{s}_z^b = \mathbf{u}_y^b = \mathbf{0}$ . The discretized Maxwell equations for the contrast field (4.60) and background field (4.63) reduce respectively to

$$-k_0^{-1} \frac{d}{dz} \mathbf{s}_{y,l}^c(z) = -\mathbf{u}_{x,l}^c(z), \quad (4.64a)$$

$$-i\mathbf{F}\mathbf{K}_x \mathbf{s}_{y,l}^c(z) = -\mathbf{u}_{z,l}^c(z), \quad (4.64b)$$

$$k_0^{-1} \frac{d}{dz} \mathbf{u}_{x,l}^c(z) + i\mathbf{F}\mathbf{K}_x \mathbf{u}_{z,l}^c(z) = -\mathbf{E}_l \mathbf{s}_{y,l}^c(z) - (\mathbf{E}_l - \mathbf{E}_l^b) \mathbf{s}_{y,l}^b(z), \quad (4.64c)$$

and

$$-k_0^{-1} \frac{d}{dz} \mathbf{s}_{y,l}^b(z) = -\mathbf{u}_{x,l}^b(z), \quad (4.65a)$$

$$-i\mathbf{F}\mathbf{K}_x \mathbf{s}_{y,l}^b(z) = -\mathbf{u}_{z,l}^b(z), \quad (4.65b)$$

$$k_0^{-1} \frac{d}{dz} \mathbf{u}_{x,l}^b(z) + i\mathbf{F}\mathbf{K}_x \mathbf{u}_{z,l}^b(z) = -\mathbf{E}_l^b \mathbf{s}_{y,l}^b(z). \quad (4.65c)$$

Substitution of (4.64a) and (4.64b) in (4.64c) yields

$$\frac{d^2}{dz^2} \mathbf{s}_{y,l}^c(z) = k_0^2 \mathbf{A}_l \mathbf{s}_{y,l}^c(z) - k_0^2 (\mathbf{E}_l - \mathbf{E}_l^b) \mathbf{s}_{y,l}^b(z), \quad (4.66)$$

where  $\mathbf{A}_l = (\mathbf{F}\mathbf{K}_x)^2 - \mathbf{E}_l$ . Equation (4.66) is a system of non-homogeneous second order ordinary differential equations. Its solution is of the form

$$\mathbf{s}_{y,l}^c = \mathbf{s}_{y,\text{hom},l}^c + \mathbf{s}_{y,\text{part},l}^c. \quad (4.67)$$



The homogeneous solution is given by

$$\mathbf{s}_{y,\text{hom},l}^c(z) = \mathbf{W}_l(e^{-k_0\mathbf{Q}_l(z-h_{l-1})}\mathbf{c}_l^+ + e^{k_0\mathbf{Q}_l(z-h_l)}\mathbf{c}_l^-), \quad (4.68)$$

where  $\mathbf{W}_l$  is the matrix of eigenvectors of  $\mathbf{A}_l$ , and  $\mathbf{Q}_l$  is a diagonal matrix with square roots of the corresponding eigenvalues on its diagonal. We assume the following form for the particular solution (method of undetermined coefficients for systems, see Chapter 3 and [59])

$$\mathbf{s}_{y,\text{part},l}^c(z) = \mathbf{p}_l s_{y,l}^b(z), \quad (4.69)$$

where  $\mathbf{p}_l \in \mathbb{R}^{2N+1}$  is a vector to be determined. Substitution of (4.69) in (4.66) yields

$$(\mathbf{A}_l - q_l^2 \mathbf{I})\mathbf{p}_l = (\mathbf{E}_l - \mathbf{E}_l^b)\mathbf{d}_0. \quad (4.70)$$

At the interface, continuity of the tangential components of the fields is required,

$$\mathbf{s}_{y,l}^c(h_l) = \mathbf{s}_{y,l+1}^c(h_l), \quad (4.71a)$$

$$\mathbf{u}_{x,l}^c(h_l) = \mathbf{u}_{x,l+1}^c(h_l). \quad (4.71b)$$

These conditions hold for the contrast field as a result of the continuity of tangential components of the total and background fields. Using (4.64a) yields

$$\mathbf{s}_{y,l}^c(h_l) = \mathbf{s}_{y,l+1}^c(h_l), \quad (4.72a)$$

$$k_0^{-1} \frac{d}{dz} \mathbf{s}_{y,l}^c(h_l) = k_0^{-1} \frac{d}{dz} \mathbf{s}_{y,l+1}^c(h_l). \quad (4.72b)$$

We define

$$\mathbf{X}_l = e^{-k_0\mathbf{Q}_l(h_l-h_{l-1})}, \quad (4.73a)$$

$$\mathbf{V}_l = -\mathbf{W}_l\mathbf{Q}_l. \quad (4.73b)$$

Then, from (4.72), (4.67), (4.68) and (4.69) we have for each slice

$$\begin{bmatrix} \mathbf{W}_l\mathbf{X}_l & \mathbf{W}_l \\ \mathbf{V}_l\mathbf{X}_l & -\mathbf{V}_l \end{bmatrix} \begin{bmatrix} \mathbf{c}_l^+ \\ \mathbf{c}_l^- \end{bmatrix} + \mathbf{g}_l(h_l) = \begin{bmatrix} \mathbf{W}_{l+1} & \mathbf{W}_{l+1}\mathbf{X}_{l+1} \\ \mathbf{V}_{l+1} & -\mathbf{V}_{l+1}\mathbf{X}_{l+1} \end{bmatrix} \begin{bmatrix} \mathbf{c}_{l+1}^+ \\ \mathbf{c}_{l+1}^- \end{bmatrix} + \mathbf{g}_{l+1}(h_l), \quad (4.74)$$

where

$$\mathbf{g}_l(z) = \begin{bmatrix} \mathbf{p}_l s_{y,l}^b \\ \mathbf{p}_l k_0^{-1} \frac{d}{dz} s_{y,l}^b \end{bmatrix} \stackrel{(4.65a)}{=} \begin{bmatrix} \mathbf{p}_l s_{y,l}^b \\ \mathbf{p}_l u_{x,l}^b \end{bmatrix}. \quad (4.75)$$

The radiation condition is imposed by requiring that coefficients of the incoming waves in layers 1 and  $M$  vanish,

$$\mathbf{c}_1^+ = \mathbf{0}, \quad \mathbf{c}_M^- = \mathbf{0}. \quad (4.76)$$

### 4.2.2 TM-polarization

In the case of planar incidence and TM-polarization we have  $\mathbf{u}_x^c = \mathbf{u}_z^c = \mathbf{s}_y^c = \mathbf{0}$  and  $\mathbf{u}_x^b = \mathbf{u}_z^b = \mathbf{s}_y^b = \mathbf{0}$ . The discretized Maxwell equations for the contrast field (4.60) and background field (4.63) reduce respectively to

$$k_0^{-1} \frac{d}{dz} \mathbf{s}_{x,l}^c(z) + i\mathbf{FK}_x \mathbf{s}_{z,l}^c(z) = -\mathbf{u}_{y,l}^c(z), \quad (4.77a)$$

$$-k_0^{-1} \frac{d}{dz} \mathbf{u}_{y,l}^c(z) = -\mathbf{P}_l^{-1} \mathbf{s}_{x,l}^c(z) - (\mathbf{P}_l^{-1} - (\mathbf{P}_l^b)^{-1}) \mathbf{s}_{x,l}^b(z), \quad (4.77b)$$

$$-i\mathbf{FK}_x \mathbf{u}_{y,l}^c(z) = -\mathbf{E}_l \mathbf{s}_{z,l}^c(z) - (\mathbf{E}_l - \mathbf{E}_l^b) \mathbf{s}_{z,l}^b(z), \quad (4.77c)$$

and

$$k_0^{-1} \frac{d}{dz} \mathbf{s}_{x,l}^b(z) + i\mathbf{FK}_x \mathbf{s}_{z,l}^b(z) = -\mathbf{u}_{y,l}^b(z), \quad (4.78a)$$

$$-k_0^{-1} \frac{d}{dz} \mathbf{u}_{y,l}^b(z) = -(\mathbf{P}_l^b)^{-1} \mathbf{s}_{x,l}^b(z), \quad (4.78b)$$

$$-i\mathbf{FK}_x \mathbf{u}_{y,l}^b(z) = -\mathbf{E}_l^b \mathbf{s}_{z,l}^b(z). \quad (4.78c)$$

A single second-order equation is obtained by substituting (4.77b) and (4.77c) in (4.77a) and subsequently using the Maxwell equations for the background field (4.78) to replace  $\mathbf{s}_{x,l}^b$  and  $\mathbf{s}_{z,l}^b$  by  $\mathbf{u}_{y,l}^b$ ,

$$\frac{d^2}{dz^2} \mathbf{u}_{y,l}^c = \mathbf{P}_l^{-1} \mathbf{B}_l \mathbf{u}_{y,l}^c + (\mathbf{P}_l^{-1} \mathbf{B}_l - (\mathbf{P}_l^b)^{-1} \mathbf{B}_l^b) \mathbf{u}_{y,l}^b, \quad (4.79)$$

where  $\mathbf{B}_l = \mathbf{FK}_x \mathbf{E}_l^{-1} \mathbf{FK}_x - \mathbf{I}$  and  $\mathbf{B}_l^b = \mathbf{FK}_x (\mathbf{E}_l^b)^{-1} \mathbf{FK}_x - \mathbf{I}$ . The solution of (4.79) is of the form

$$\mathbf{u}_{y,l}^c = \mathbf{u}_{y,\text{hom},l}^c + \mathbf{u}_{y,\text{part},l}^c. \quad (4.80)$$

The homogeneous solution is given by

$$\mathbf{u}_{y,\text{hom},l}^c(z) = \mathbf{W}_l (e^{-k_0 \mathbf{Q}_l (z-h_{l-1})} \mathbf{c}_l^+ + e^{k_0 \mathbf{Q}_l (z-h_l)} \mathbf{c}_l^-), \quad (4.81)$$

where, as before,  $\mathbf{W}_l$  and  $\mathbf{Q}_l$  are respectively the matrix of eigenvectors and diagonal matrix of square roots of eigenvalues of  $\mathbf{P}_l^{-1} \mathbf{B}_l$ . We assume the following form for the

particular solution

$$\mathbf{u}_{y,\text{part},l}^c(z) = \mathbf{p}_l u_{y,l}^b(z). \quad (4.82)$$

Substitution of (4.82) in (4.79) yields

$$(\mathbf{P}_l^{-1} \mathbf{B}_l - q_l^2 \mathbf{I}) \mathbf{p}_l = ((\mathbf{P}_l^b)^{-1} \mathbf{B}_l^b - \mathbf{P}_l^{-1} \mathbf{B}_l) \mathbf{d}_0. \quad (4.83)$$

At the interface, continuity of the tangential components of the fields is required,

$$\mathbf{u}_{y,l}^c(h_l) = \mathbf{u}_{y,l+1}^c(h_l), \quad (4.84a)$$

$$\mathbf{s}_{x,l}^c(h_l) = \mathbf{s}_{x,l+1}^c(h_l). \quad (4.84b)$$

Using (4.77b) yields

$$\mathbf{u}_{y,l}^c(h_l) = \mathbf{u}_{y,l+1}^c(h_l), \quad (4.85a)$$

$$\begin{aligned} k_0^{-1} \mathbf{P}_l \frac{d}{dz} \mathbf{u}_{y,l}^c(h_l) - \mathbf{P}_l (\mathbf{P}_l^{-1} - (\mathbf{P}_l^b)^{-1}) \mathbf{s}_{x,l}^b(h_l) = \\ k_0^{-1} \mathbf{P}_{l+1} \frac{d}{dz} \mathbf{u}_{y,l+1}^c(h_l) - \mathbf{P}_{l+1} (\mathbf{P}_{l+1}^{-1} - (\mathbf{P}_{l+1}^b)^{-1}) \mathbf{s}_{x,l+1}^b(h_l). \end{aligned} \quad (4.85b)$$

We define  $\mathbf{V}_l = -\mathbf{P}_l \mathbf{W}_l \mathbf{Q}_l$  (keeping the old definition of  $\mathbf{X}_l$ ). Then, from (4.85), (4.80), (4.81) and (4.82) we have for each slice

$$\begin{bmatrix} \mathbf{W}_l \mathbf{X}_l & \mathbf{W}_l \\ \mathbf{V}_l \mathbf{X}_l & -\mathbf{V}_l \end{bmatrix} \begin{bmatrix} \mathbf{c}_l^+ \\ \mathbf{c}_l^- \end{bmatrix} + \mathbf{g}_l(h_l) = \begin{bmatrix} \mathbf{W}_{l+1} & \mathbf{W}_{l+1} \mathbf{X}_{l+1} \\ \mathbf{V}_{l+1} & -\mathbf{V}_{l+1} \mathbf{X}_{l+1} \end{bmatrix} \begin{bmatrix} \mathbf{c}_{l+1}^+ \\ \mathbf{c}_{l+1}^- \end{bmatrix} + \mathbf{g}_{l+1}(h_l), \quad (4.86)$$

where

$$\begin{aligned} \mathbf{g}_l(z) = & \begin{bmatrix} \mathbf{P}_l u_{y,l}^b \\ \mathbf{P}_l \mathbf{p}_l k_0^{-1} \frac{d}{dz} u_{y,l}^b - \mathbf{P}_l (\mathbf{P}_l^{-1} - (\mathbf{P}_l^b)^{-1}) \mathbf{d}_0 s_{x,l}^b \end{bmatrix} \\ \stackrel{(4.78b)}{=} & \begin{bmatrix} \mathbf{P}_l u_{y,l}^b \\ (\mathbf{P}_l (\mathbf{P}_l^b)^{-1} \mathbf{p}_l + (\mathbf{P}_l - \mathbf{P}_l^b) (\mathbf{P}_l^b)^{-1} \mathbf{d}_0) s_{x,l}^b \end{bmatrix}. \end{aligned} \quad (4.87)$$

Similarly to the TE-case, the radiation condition is imposed by requiring that coefficients of the incoming waves in layers 1 and  $M$  vanish,

$$\mathbf{c}_1^+ = \mathbf{0}, \quad \mathbf{c}_M^- = \mathbf{0}. \quad (4.88)$$

### 4.2.3 Conical incidence

In the case of conical incidence, the Maxwell equations for the contrast field (4.60) reduce to

$$\frac{d^2}{dz^2} \mathbf{s}_{x,l}^c(z) = k_0^2 \mathbf{C}_l \mathbf{s}_{x,l}^c(z) + k_0^2 (\mathbf{B}_l \mathbf{P}_l^{-1} - \mathbf{B}_l^b (\mathbf{P}_l^b)^{-1}) \mathbf{s}_{x,l}^b(z), \quad (4.89a)$$

$$\frac{d^2}{dz^2} \mathbf{u}_{x,l}^c(z) = k_0^2 \mathbf{D}_l \mathbf{u}_{x,l}^c(z) - k_0^2 (\mathbf{E}_l - \mathbf{E}_l^b) \mathbf{u}_{x,l}^b(z), \quad (4.89b)$$

where  $\mathbf{C}_l = \mathbf{K}_y^2 + \mathbf{B}_l \mathbf{P}_l^{-1}$ ,  $\mathbf{D}_l = \mathbf{K}_y^2 + \mathbf{A}_l$ . As for the TM-polarization, in order to arrive at (4.89) from (4.60), additionally the Maxwell equations for the background field (4.63) need to be used. The solution vector is of the form

$$\mathbf{s}_{x,l}^c = \mathbf{s}_{x,\text{hom},l}^c + \mathbf{s}_{x,\text{part},l}^c, \quad (4.90a)$$

$$\mathbf{u}_{x,l}^c = \mathbf{u}_{x,\text{hom},l}^c + \mathbf{u}_{x,\text{part},l}^c. \quad (4.90b)$$

The homogeneous solution is given by

$$\mathbf{s}_{x,\text{hom},l}^c(z) = \mathbf{W}_{s,l} (e^{-k_0 \mathbf{Q}_{s,l} (z-h_{l-1})} \mathbf{c}_{s,l}^+ + e^{k_0 \mathbf{Q}_{s,l} (z-h_l)} \mathbf{c}_{s,l}^-), \quad (4.91)$$

$$\mathbf{u}_{x,\text{hom},l}^c(z) = \mathbf{W}_{u,l} (e^{-k_0 \mathbf{Q}_{u,l} (z-h_{l-1})} \mathbf{c}_{u,l}^+ + e^{k_0 \mathbf{Q}_{u,l} (z-h_l)} \mathbf{c}_{u,l}^-), \quad (4.92)$$

where the pairs  $\mathbf{W}_{s,l}$ ,  $\mathbf{Q}_{s,l}$  and  $\mathbf{W}_{u,l}$ ,  $\mathbf{Q}_{u,l}$  contain the matrix of eigenvectors and the diagonal matrix of square roots of eigenvalues of, respectively,  $\mathbf{C}_l$  and  $\mathbf{D}_l$ . To find the particular solution we assume the form

$$\mathbf{s}_{x,\text{part},l}^c(z) = \mathbf{p}_{s,l} s_{x,l}^b(z), \quad (4.93a)$$

$$\mathbf{u}_{x,\text{part},l}^c(z) = \mathbf{p}_{u,l} u_{x,l}^b(z). \quad (4.93b)$$

Using this Ansatz in Equation (4.89), we obtain two linear systems that can be solved for  $\mathbf{p}_{s,l}$  and  $\mathbf{p}_{u,l}$

$$(\mathbf{C}_l - q_l^2 \mathbf{I}) \mathbf{p}_{s,l} = -(\mathbf{B}_l \mathbf{P}_l^{-1} - \mathbf{B}_l^b (\mathbf{P}_l^b)^{-1}) \mathbf{d}_0, \quad (4.94a)$$

$$(\mathbf{D}_l - q_l^2 \mathbf{I}) \mathbf{p}_{u,l} = (\mathbf{E}_l - \mathbf{E}_l^b) \mathbf{d}_0. \quad (4.94b)$$

At the interface, continuity of the tangential components of the fields is required

$$\mathbf{s}_{x,l}^c(h_l) = \mathbf{s}_{x,l+1}^c(h_l), \quad (4.95a)$$

$$\mathbf{s}_{y,l}^c(h_l) = \mathbf{s}_{y,l+1}^c(h_l), \quad (4.95b)$$

$$\mathbf{u}_{x,l}^c(h_l) = \mathbf{u}_{x,l+1}^c(h_l), \quad (4.95c)$$

$$\mathbf{u}_{y,l}^c(h_l) = \mathbf{u}_{y,l+1}^c(h_l). \quad (4.95d)$$

Using (4.60c) in (4.60e) and (4.60f) in (4.60b), the  $y$ -components of the fields are expressed in terms of  $x$ -components,

$$\mathbf{s}_{y,l}^c = \mathbf{A}_l^{-1} \left( \mathbf{F}\mathbf{K}_x\mathbf{K}_y\mathbf{s}_{x,l}^c + k_0^{-1} \frac{d}{dz} \mathbf{u}_{x,l}^c + (\mathbf{E}_l - \mathbf{E}_l^b)\mathbf{s}_{y,l}^b \right), \quad (4.96a)$$

$$\mathbf{u}_{y,l}^c = \mathbf{B}_l^{-1} \left( \mathbf{F}\mathbf{K}_x\mathbf{E}_l^{-1}\mathbf{K}_y\mathbf{u}_{x,l}^c + k_0^{-1} \frac{d}{dz} \mathbf{s}_{x,l}^c - i\mathbf{F}\mathbf{K}_x(\mathbf{I} - \mathbf{E}_l^b\mathbf{E}_l^{-1})\mathbf{s}_{z,l}^b \right). \quad (4.96b)$$

We define

$$\mathbf{W}_l = \begin{bmatrix} \mathbf{0} & \mathbf{W}_{s,l} \\ \mathbf{A}_l^{-1}\mathbf{W}_{u,l}\mathbf{Q}_{u,l} & \mathbf{A}_l^{-1}\mathbf{F}\mathbf{K}_x\mathbf{K}_y\mathbf{W}_{s,l} \end{bmatrix}, \quad (4.97a)$$

$$\mathbf{V}_l = - \begin{bmatrix} \mathbf{W}_{u,l} & \mathbf{0} \\ \mathbf{B}_l^{-1}\mathbf{F}\mathbf{K}_x\mathbf{E}_l^{-1}\mathbf{K}_y\mathbf{W}_{u,l} & \mathbf{B}_l^{-1}\mathbf{W}_{s,l}\mathbf{Q}_{s,l} \end{bmatrix}, \quad (4.97b)$$

and

$$\mathbf{c}_l^+ = \begin{bmatrix} -\mathbf{c}_{u,l}^+ \\ \mathbf{c}_{s,l}^+ \end{bmatrix}, \quad \mathbf{c}_l^- = \begin{bmatrix} \mathbf{c}_{u,l}^- \\ \mathbf{c}_{s,l}^- \end{bmatrix}. \quad (4.97c)$$

Then, from (4.95), (4.90), (4.91) and (4.93) we have for each slice

$$\begin{bmatrix} \mathbf{W}_l\mathbf{X}_l & \mathbf{W}_l \\ \mathbf{V}_l\mathbf{X}_l & -\mathbf{V}_l \end{bmatrix} \begin{bmatrix} \mathbf{c}_l^+ \\ \mathbf{c}_l^- \end{bmatrix} + \mathbf{g}_l(h_l) = \begin{bmatrix} \mathbf{W}_{l+1} & \mathbf{W}_{l+1}\mathbf{X}_{l+1} \\ \mathbf{V}_{l+1} & -\mathbf{V}_{l+1}\mathbf{X}_{l+1} \end{bmatrix} \begin{bmatrix} \mathbf{c}_{l+1}^+ \\ \mathbf{c}_{l+1}^- \end{bmatrix} + \mathbf{g}_{l+1}(h_l), \quad (4.98)$$

where

$$\mathbf{g}_l(z) = \begin{bmatrix} \mathbf{A}_l^{-1}(\mathbf{F}\mathbf{K}_x\mathbf{K}_y\mathbf{p}_{s,l}^b s_{x,l}^b + k_0^{-1}\mathbf{p}_{u,l}^b \frac{d}{dz} u_{x,l}^b + (\mathbf{E}_l - \mathbf{E}_l^b)\mathbf{d}_0 s_{y,l}^b) \\ \mathbf{B}_l^{-1}(\mathbf{F}\mathbf{K}_x\mathbf{E}_l^{-1}\mathbf{K}_y\mathbf{p}_{u,l}^b u_{x,l}^b + k_0^{-1}\mathbf{p}_{s,l}^b \frac{d}{dz} s_{x,l}^b - i\mathbf{F}\mathbf{K}_x(\mathbf{I} - \mathbf{E}_l^b\mathbf{E}_l^{-1})\mathbf{d}_0 s_{z,l}^b) \end{bmatrix},$$

with  $s_{z,l}^b = \frac{i}{\epsilon^b}(k_{x0}u_{y,l}^b - k_y u_{x,l}^b)$ . The radiation condition is imposed by requiring that coefficients of the incoming waves in layers 1 and  $M$  vanish,

$$\mathbf{c}_1^+ = \mathbf{0}, \quad \mathbf{c}_M^- = \mathbf{0}. \quad (4.99)$$

### 4.3 Final remarks

In this chapter the FMM and the AFMM-CFF have been generalized to arbitrary shapes and illumination. It has been shown that in the FMM discretization leads to homogeneous linear systems, while in the AFMM-CFF the source term causes the resulting linear systems to become non-homogeneous. When the radiation conditions are used, these systems have the same number of equations and unknowns. Due to stability issues

special care needs to be taken when solving the coupled linear systems. This is discussed in Chapter 5. For this reason presentation of numerical results is postponed to the next chapter.



## Chapter 5

# Stable solution of the coupled linear systems

In order to apply the FMM to non-periodic structures, perfectly matched layers need to be placed at the periodic boundaries and the Maxwell equations have to be formulated in terms of a contrast (scattered) field. This reformulation modifies the structure of the resulting linear systems and makes the application of available stable recursion algorithms impossible. We adapt the well-known S-matrix algorithm for use with the aperiodic Fourier modal method in contrast field formulation (AFMM-CFF). To this end, stable recursive relations are derived for linear systems with non-homogeneous structure. The stability of the algorithm is confirmed by numerical results.

### 5.1 Introduction

The classical Fourier modal method (FMM) [47, 65] and the aperiodic Fourier modal method in contrast-field formulation (AFMM-CFF) (see Chapters 3 and 4 as well as [59, 58]) rely on two main steps: {1} discretizing the computational domain into slices and obtaining the general solution (up to integration constants) in each slice and {2} determining the integration constants (or modal field amplitudes) by solving a sequence of recursive linear systems resulting from the application of the interface conditions. The straightforward approach for solving the sequence of linear systems, the T-matrix algorithm, is known to be numerically unstable [39]. This issue is common for various numerical methods in optics and electromagnetics when discretizing the direction normal to the layered media and is generally linked to the growing exponentials appearing in the equations.



During the past two decades many algorithms for solving the sequence of recursive linear systems that avoid instability issues have been proposed [55, 17, 39, 9]. Many of them are, in one way or another, connected to the *S-matrix algorithm* [32]. Even for the *enhanced transmittance matrix approach* [48], which has a somewhat different algebraic structure, the connection to the S-matrix algorithm has been revealed [84]. A recent study [88] shows that the enhanced transmittance matrix approach is equivalent to a stable condensation algorithm based on Riccati transformations. For a stability study of the latter see [1] and references therein. We refer the reader to Ref. [40] for an account of the S-matrix, and related *R-matrix algorithms*. It is worth mentioning that an efficient parallel implementation of the S-matrix algorithm has been recently presented [30].

The S-matrix algorithm relies on the physical concept of mapping the incoming waves on an interface to outgoing waves. This mapping is realized by a so-called S-matrix. Therefore the S-matrix algorithm is suited for linear systems with a homogeneous structure (of the type  $\mathbf{A}_l \mathbf{x}_l = \mathbf{A}_{l+1} \mathbf{x}_{l+1}$ , as opposed to the more general non-homogeneous case  $\mathbf{A}_l \mathbf{x}_l + \mathbf{f}_l = \mathbf{A}_{l+1} \mathbf{x}_{l+1} + \mathbf{f}_{l+1}$ ). In the AFMM-CFF, the modification of the interface conditions and of the general solution leads to a non-homogeneous structure of the linear systems. We adapt the existing S-matrix algorithm to the new structure of the equations. The choice of the S-matrix approach over the enhanced transmittance matrix approach as a starting point for the extension is explained by the superior flexibility and generality of the former.

This chapter is structured as follows. In Section 5.2 we present classical T-matrix and S-matrix algorithms for homogeneous linear systems. In Section 5.3 we demonstrate how the classical S-matrix algorithm can be adapted to the non-homogeneous linear systems arising in the AFMM-CFF. Section 5.4 contains numerical results confirming the stability of the proposed method. The problem of scattering from a dielectric cylinder, which admits a semi-analytical solution, is used for this purpose.

## 5.2 Homogeneous T-matrix and S-matrix algorithms

In this section we are concerned with solving the homogeneous recursive linear systems (4.17), (4.27), (4.46) arising in the standard FMM. It has been shown in Section 4.1 that for all cases (TE, TM and conical) these systems are of the form

$$\mathbf{R}_l \begin{bmatrix} \mathbf{X}_l & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{c}_l^+ \\ \mathbf{c}_l^- \end{bmatrix} = \mathbf{R}_{l+1} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_{l+1} \end{bmatrix} \begin{bmatrix} \mathbf{c}_{l+1}^+ \\ \mathbf{c}_{l+1}^- \end{bmatrix}, \quad (5.1)$$

where

$$\mathbf{R}_l = \begin{bmatrix} \mathbf{W}_l & \mathbf{W}_l \\ \mathbf{V}_l & \mathbf{V}_l \end{bmatrix}. \quad (5.2)$$

A straightforward approach of solving this system is to eliminate the coefficients in the intermediary slices. For this purpose the interface relation (5.1) can be written as a T-matrix interface relation:

$$\begin{bmatrix} \mathbf{c}_{l+1}^+ \\ \mathbf{c}_{l+1}^- \end{bmatrix} = \begin{bmatrix} \mathbf{T}_l^{11} & \mathbf{T}_l^{12} \\ \mathbf{T}_l^{21} & \mathbf{T}_l^{22} \end{bmatrix} \begin{bmatrix} \mathbf{c}_l^+ \\ \mathbf{c}_l^- \end{bmatrix}, \quad (5.3)$$

where

$$\mathbf{T}_l = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_{l+1} \end{bmatrix}^{-1} \mathbf{T}'_l \begin{bmatrix} \mathbf{X}_l & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}, \quad (5.4)$$

with

$$\mathbf{T}'_l = \mathbf{R}_{l+1}^{-1} \mathbf{R}_l = \frac{1}{2} \begin{bmatrix} \mathbf{W}_{l+1}^{-1} & \mathbf{V}_{l+1}^{-1} \\ \mathbf{W}_{l+1}^{-1} & -\mathbf{V}_{l+1}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{W}_l & \mathbf{W}_l \\ \mathbf{V}_l & -\mathbf{V}_l \end{bmatrix}. \quad (5.5)$$

Elimination of the intermediary coefficients yields

$$\begin{bmatrix} \mathbf{c}_M^+ \\ \mathbf{c}_M^- \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{T}}_l^{11} & \bar{\mathbf{T}}_l^{12} \\ \bar{\mathbf{T}}_l^{21} & \bar{\mathbf{T}}_l^{22} \end{bmatrix} \begin{bmatrix} \mathbf{c}_1^+ \\ \mathbf{c}_1^- \end{bmatrix}, \quad (5.6)$$

where

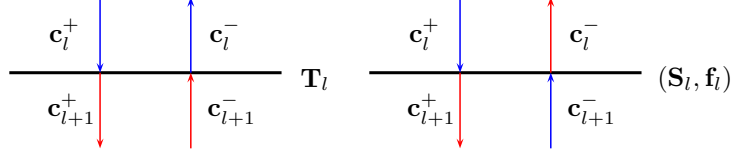
$$\bar{\mathbf{T}}_{M-1} = \prod_{l=M-1}^1 \mathbf{T}_l. \quad (5.7)$$

Since the vector coefficients  $\mathbf{c}_M^-$  and  $\mathbf{c}_1^-$  are known from the radiation conditions (see Chapter 4), the system (5.6) has two vector equations with two vector unknowns and can be easily solved. If needed, the intermediary coefficients are determined at a later stage. Relations (5.3), (5.4), (5.6) and (5.7) define the so-called *T*-matrix algorithm, which has a very simple implementation. Unfortunately this algorithm is unstable. The reason of instability lies in the matrix

$$\mathbf{X}_{l+1} = e^{-k_0 \mathbf{Q}_{l+1} (h_{l+2} - h_{l+1})}. \quad (5.8)$$

For thick slices or a large number of harmonics some entries of  $\mathbf{X}_{l+1}$  become extremely small. The entries of  $\mathbf{X}_{l+1}^{-1}$  computed in (5.4) consequently become extremely large and introduce a large round-off error when represented in floating point arithmetic.

We now discuss a stable alternative to the T-matrix algorithm: the *S*-matrix algorithm.



**Figure 5.1:** *T*-matrix (left) and *S*-matrix (right) representation of an interface relation. Colors represent the input-output properties of the *T*-matrix and *S*-matrix: blue waves are mapped to red waves.

Unlike in the *T*-matrix representation of an interface relation, an *S*-matrix maps incoming waves (on the interface) to outgoing waves (see Figure 5.1):

$$\begin{bmatrix} \mathbf{c}_{l+1}^+ \\ \mathbf{c}_l^- \end{bmatrix} = \begin{bmatrix} \mathbf{S}_l^{11} & \mathbf{S}_l^{12} \\ \mathbf{S}_l^{21} & \mathbf{S}_l^{22} \end{bmatrix} \begin{bmatrix} \mathbf{c}_l^+ \\ \mathbf{c}_{l+1}^- \end{bmatrix}. \quad (5.9)$$

The *S*-matrix  $\mathbf{S}_l$  of an interface is related to the *T*-matrix  $\mathbf{T}_l$  of the same interface by

$$\mathbf{S}_l^{11} = (\mathbf{T}_l^{\prime 11} - \mathbf{T}_l^{\prime 12} (\mathbf{T}_l^{\prime 22})^{-1} \mathbf{T}_l^{\prime 21}) \mathbf{X}_l, \quad (5.10a)$$

$$\mathbf{S}_l^{12} = \mathbf{T}_l^{\prime 12} (\mathbf{T}_l^{\prime 22})^{-1} \mathbf{X}_{l+1}, \quad (5.10b)$$

$$\mathbf{S}_l^{21} = -(\mathbf{T}_l^{\prime 22})^{-1} \mathbf{T}_l^{\prime 21} \mathbf{X}_l, \quad (5.10c)$$

$$\mathbf{S}_l^{22} = (\mathbf{T}_l^{\prime 22})^{-1} \mathbf{X}_{l+1}. \quad (5.10d)$$

The above relations are obtained by bringing (5.3) to form (5.9). Similarly to the *T*-matrix algorithm, the intermediary coefficients will be eliminated such that an expression of the following form is obtained for the coefficients in the layers 1 and  $M$ :

$$\begin{bmatrix} \mathbf{c}_M^+ \\ \mathbf{c}_1^- \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{S}}_{M-1}^{11} & \bar{\mathbf{S}}_{M-1}^{12} \\ \bar{\mathbf{S}}_{M-1}^{21} & \bar{\mathbf{S}}_{M-1}^{22} \end{bmatrix} \begin{bmatrix} \mathbf{c}_1^+ \\ \mathbf{c}_M^- \end{bmatrix}. \quad (5.11)$$

As shown in [32, 40] the *cumulative scattering matrix*  $\bar{\mathbf{S}}_{M-1}$  can be computed recursively in a stable manner using the update relation

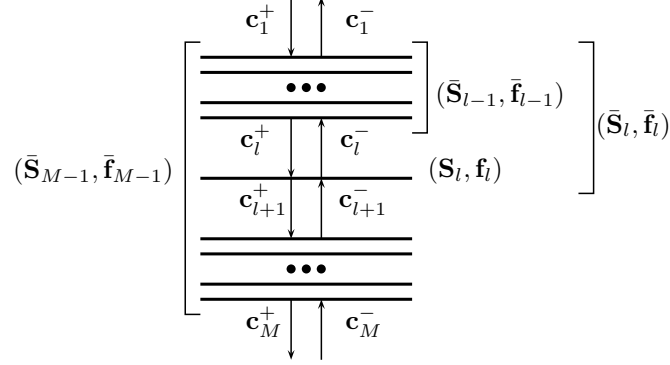
$$\begin{bmatrix} \bar{\mathbf{S}}_l^{11} & \bar{\mathbf{S}}_l^{12} \\ \bar{\mathbf{S}}_l^{21} & \bar{\mathbf{S}}_l^{22} \end{bmatrix} = \begin{bmatrix} \mathbf{S}_l^{11} \mathbf{H}'_l \bar{\mathbf{S}}_{l-1}^{11} & \mathbf{S}_l^{12} + \mathbf{S}_l^{11} \bar{\mathbf{S}}_{l-1}^{12} \mathbf{H}''_l \mathbf{S}_l^{22} \\ \bar{\mathbf{S}}_{l-1}^{21} + \bar{\mathbf{S}}_{l-1}^{22} \mathbf{S}_l^{21} \mathbf{H}'_l \bar{\mathbf{S}}_{l-1}^{11} & \bar{\mathbf{S}}_{l-1}^{22} \mathbf{H}''_l \mathbf{S}_l^{22} \end{bmatrix}, \quad (5.12)$$

with

$$\mathbf{H}'_l = (\mathbf{I} - \bar{\mathbf{S}}_{l-1}^{12} \mathbf{S}_l^{21})^{-1}, \quad (5.13)$$

$$\mathbf{H}''_l = (\mathbf{I} - \mathbf{S}_l^{21} \bar{\mathbf{S}}_{l-1}^{12})^{-1}. \quad (5.14)$$

Note that the *T*-matrix algorithm can also be seen as a recursive algorithm with an



**Figure 5.2:** A stack of interfaces with upward- and downward-traveling waves in-between.

update relation

$$\bar{\mathbf{T}}_l = \mathbf{T}_l \bar{\mathbf{T}}_{l-1}. \quad (5.15)$$

### 5.3 Non-homogeneous S-matrix algorithm

We now proceed to adapt the standard S-matrix algorithm for use with non-homogeneous recursive linear systems arising in the AFMM-CFF. As shown in Section 4.2, matching of the interface conditions for the contrast field at interface  $l$  yields an equation of the form

$$\mathbf{R}_l \begin{bmatrix} \mathbf{X}_l & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{c}_l^+ \\ \mathbf{c}_l^- \end{bmatrix} + \mathbf{g}_l(h_{l+1}) = \mathbf{R}_{l+1} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_{l+1} \end{bmatrix} \begin{bmatrix} \mathbf{c}_{l+1}^+ \\ \mathbf{c}_{l+1}^- \end{bmatrix} + \mathbf{g}_{l+1}(h_{l+1}). \quad (5.16)$$

This is rewritten in the T-matrix formalism as

$$\begin{bmatrix} \mathbf{c}_{l+1}^+ \\ \mathbf{c}_{l+1}^- \end{bmatrix} = \begin{bmatrix} \mathbf{T}_l^{11} & \mathbf{T}_l^{12} \\ \mathbf{T}_l^{21} & \mathbf{T}_l^{22} \end{bmatrix} \begin{bmatrix} \mathbf{c}_l^+ \\ \mathbf{c}_l^- \end{bmatrix} + \begin{bmatrix} \mathbf{g}'_l{}^1 \\ \mathbf{g}'_l{}^2 \end{bmatrix}, \quad (5.17)$$

where  $\mathbf{T}_l$  is defined in (5.4) and

$$\mathbf{g}'_l = \mathbf{R}_{l+1}^{-1}(\mathbf{g}_l(h_{l+1}) - \mathbf{g}_{l+1}(h_{l+1})). \quad (5.18)$$

The S-matrix algorithm is derived from the matrix  $\mathbf{T}_l$  and vector  $\mathbf{g}'_l$  of the T-matrix equation (5.17). In the spirit of the S-matrix algorithm, the waves scattered at the interface are expressed in terms of waves incident on the interface [40] (see Figure 5.2)

$$\begin{bmatrix} \mathbf{c}_{l+1}^+ \\ \mathbf{c}_l^- \end{bmatrix} = \begin{bmatrix} \mathbf{S}_l^{11} & \mathbf{S}_l^{12} \\ \mathbf{S}_l^{21} & \mathbf{S}_l^{22} \end{bmatrix} \begin{bmatrix} \mathbf{c}_l^+ \\ \mathbf{c}_{l+1}^- \end{bmatrix} + \begin{bmatrix} \mathbf{f}_l^1 \\ \mathbf{f}_l^2 \end{bmatrix}. \quad (5.19)$$

The S-matrix  $\mathbf{S}_l$  and the vector  $\mathbf{f}_l$  can be determined from the T-matrix  $\mathbf{T}_l$  and the vector  $\mathbf{g}_l'$  by bringing (5.17) to form (5.19).

$$\mathbf{S}_l^{11} = (\mathbf{T}_l'^{11} - \mathbf{T}_l'^{12}(\mathbf{T}_l'^{22})^{-1}\mathbf{T}_l'^{21})\mathbf{X}_l, \quad (5.20a)$$

$$\mathbf{S}_l^{12} = \mathbf{T}_l'^{12}(\mathbf{T}_l'^{22})^{-1}\mathbf{X}_{l+1}, \quad (5.20b)$$

$$\mathbf{S}_l^{21} = -(\mathbf{T}_l'^{22})^{-1}\mathbf{T}_l'^{21}\mathbf{X}_l, \quad (5.20c)$$

$$\mathbf{S}_l^{22} = (\mathbf{T}_l'^{22})^{-1}\mathbf{X}_{l+1}, \quad (5.20d)$$

$$\mathbf{f}_l^1 = \mathbf{g}_l'^1 - \mathbf{T}_l'^{12}(\mathbf{T}_l'^{22})^{-1}\mathbf{g}_l'^2, \quad (5.20e)$$

$$\mathbf{f}_l^2 = -(\mathbf{T}_l'^{22})^{-1}\mathbf{g}_l'^2. \quad (5.20f)$$

Expressions (5.19) and (5.20) describe the scattering properties of the interface  $l$ . These properties are defined by a *local scattering matrix*  $\mathbf{S}_l$  and a *local source vector*  $\mathbf{f}_l$ . To simplify further presentation we denote them together as  $(\mathbf{S}_l, \mathbf{f}_l)$ . Figure 5.2 gives a schematic representation of the interfaces and associated scattering matrix-vector pairs.

We proceed by defining a *cumulative scattering matrix* and a *cumulative source vector* for a stack of multiple interfaces. The matrix-vector pair  $(\bar{\mathbf{S}}_l, \bar{\mathbf{f}}_l)$  defines the scattering properties of the stack of interfaces  $1, \dots, l$  (see also Figure 5.2).

$$\begin{bmatrix} \mathbf{c}_{l+1}^+ \\ \mathbf{c}_1^- \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{S}}_l^{11} & \bar{\mathbf{S}}_l^{12} \\ \bar{\mathbf{S}}_l^{21} & \bar{\mathbf{S}}_l^{22} \end{bmatrix} \begin{bmatrix} \mathbf{c}_1^+ \\ \mathbf{c}_{l+1}^- \end{bmatrix} + \begin{bmatrix} \bar{\mathbf{f}}_l^1 \\ \bar{\mathbf{f}}_l^2 \end{bmatrix}. \quad (5.21)$$

By assuming that the cumulative scattering matrix-vector pair for interface  $l-1$  is known, we will derive the cumulative matrix-vector pair for interface  $l$  using the local scattering matrix-vector of interface  $l$ , as illustrated by the diagram

$$(\bar{\mathbf{S}}_{l-1}, \bar{\mathbf{f}}_{l-1}) \xrightarrow{(\mathbf{S}_l, \mathbf{f}_l)} (\bar{\mathbf{S}}_l, \bar{\mathbf{f}}_l). \quad (5.22)$$

This defines a recursive relation for the cumulative matrix-vector. Note that the cumulative scattering matrix-vector and the local scattering matrix-vector for interface one are equal,

$$(\bar{\mathbf{S}}_1, \bar{\mathbf{f}}_1) = (\mathbf{S}_1, \mathbf{f}_1). \quad (5.23)$$

This relation is used to initialize the recursion.

We now outline the derivation of the recursion formally represented in (5.22). For con-

venience, relation (5.21) is repeated here for interface  $l - 1$

$$\begin{bmatrix} \mathbf{c}_l^+ \\ \mathbf{c}_1^- \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{S}}_{l-1}^{11} & \bar{\mathbf{S}}_{l-1}^{12} \\ \bar{\mathbf{S}}_{l-1}^{21} & \bar{\mathbf{S}}_{l-1}^{22} \end{bmatrix} \begin{bmatrix} \mathbf{c}_1^+ \\ \mathbf{c}_l^- \end{bmatrix} + \begin{bmatrix} \bar{\mathbf{f}}_{l-1}^1 \\ \bar{\mathbf{f}}_{l-1}^2 \end{bmatrix}. \quad (5.24)$$

The first equation of (5.24) and the second equation of (5.19) yield

$$\mathbf{c}_l^+ = \mathbf{H}'_l \bar{\mathbf{S}}_{l-1}^{11} \mathbf{c}_1^+ + \bar{\mathbf{S}}_{l-1}^{12} \mathbf{H}''_l \mathbf{S}_{l+1}^{22} \mathbf{c}_{l+1}^- + \mathbf{H}'_l \bar{\mathbf{S}}_{l-1}^{12} \mathbf{f}_l^2 + \mathbf{H}'_l \bar{\mathbf{f}}_{l-1}^1, \quad (5.25)$$

where

$$\mathbf{H}'_l = (\mathbf{I} - \bar{\mathbf{S}}_{l-1}^{12} \mathbf{S}_l^{21})^{-1}, \quad (5.26a)$$

$$\mathbf{H}''_l = (\mathbf{I} - \mathbf{S}_l^{21} \bar{\mathbf{S}}_{l-1}^{12})^{-1}. \quad (5.26b)$$

From the first equation of (5.19) and (5.25)

$$\begin{aligned} \mathbf{c}_{l+1}^+ &= \mathbf{S}_l^{11} \mathbf{H}'_l \bar{\mathbf{S}}_{l-1}^{11} \mathbf{c}_1^+ + (\mathbf{S}_l^{12} + \mathbf{S}_l^{11} \bar{\mathbf{S}}_{l-1}^{12} \mathbf{H}''_l \mathbf{S}_l^{22}) \mathbf{c}_{l+1}^- \\ &\quad + \mathbf{S}_l^{11} (\mathbf{H}'_l \bar{\mathbf{S}}_{l-1}^{12} \mathbf{f}_l^2 + \mathbf{H}'_l \bar{\mathbf{f}}_{l-1}^1) + \mathbf{f}_l^1. \end{aligned} \quad (5.27)$$

From the second equation of (5.24), the second equation of (5.19) and (5.25)

$$\begin{aligned} \mathbf{c}_1^- &= (\bar{\mathbf{S}}_{l-1}^{21} + \bar{\mathbf{S}}_{l-1}^{22} \mathbf{S}_l^{21} \mathbf{H}'_l \bar{\mathbf{S}}_{l-1}^{11}) \mathbf{c}_1^+ + \bar{\mathbf{S}}_{l-1}^{22} \mathbf{H}''_l \mathbf{S}_l^{22} \mathbf{c}_{l+1}^- \\ &\quad + \bar{\mathbf{S}}_{l-1}^{22} \mathbf{S}_l^{21} (\mathbf{H}'_l \bar{\mathbf{S}}_{l-1}^{12} \mathbf{f}_l^2 + \mathbf{H}'_l \bar{\mathbf{f}}_{l-1}^1) + \bar{\mathbf{S}}_{l-1}^{22} \mathbf{f}_l^2 + \bar{\mathbf{f}}_{l-1}^2. \end{aligned} \quad (5.28)$$

Now (5.27) and (5.28) give the recursion relations

$$\begin{bmatrix} \bar{\mathbf{S}}_l^{11} & \bar{\mathbf{S}}_l^{12} \\ \bar{\mathbf{S}}_l^{21} & \bar{\mathbf{S}}_l^{22} \end{bmatrix} = \begin{bmatrix} \mathbf{S}_l^{11} \mathbf{H}'_l \bar{\mathbf{S}}_{l-1}^{11} & \mathbf{S}_l^{12} + \mathbf{S}_l^{11} \bar{\mathbf{S}}_{l-1}^{12} \mathbf{H}''_l \mathbf{S}_l^{22} \\ \bar{\mathbf{S}}_{l-1}^{21} + \bar{\mathbf{S}}_{l-1}^{22} \mathbf{S}_l^{21} \mathbf{H}'_l \bar{\mathbf{S}}_{l-1}^{11} & \bar{\mathbf{S}}_{l-1}^{22} \mathbf{H}''_l \mathbf{S}_l^{22} \end{bmatrix}, \quad (5.29a)$$

$$\begin{bmatrix} \bar{\mathbf{f}}_l^1 \\ \bar{\mathbf{f}}_l^2 \end{bmatrix} = \begin{bmatrix} \mathbf{S}_l^{11} (\mathbf{H}'_l \bar{\mathbf{S}}_{l-1}^{12} \mathbf{f}_l^2 + \mathbf{H}'_l \bar{\mathbf{f}}_{l-1}^1) + \mathbf{f}_l^1 \\ \bar{\mathbf{S}}_{l-1}^{22} \mathbf{S}_l^{21} (\mathbf{H}'_l \bar{\mathbf{S}}_{l-1}^{12} \mathbf{f}_l^2 + \mathbf{H}'_l \bar{\mathbf{f}}_{l-1}^1) + \bar{\mathbf{S}}_{l-1}^{22} \mathbf{f}_l^2 + \bar{\mathbf{f}}_{l-1}^2 \end{bmatrix}. \quad (5.29b)$$

These update relations define the *non-homogeneous S-matrix algorithm*. While the update relation for the matrix  $\bar{\mathbf{S}}_l$  is the same as in the classical S-matrix algorithm (see relation (5.12) in this chapter and relation (15a) in Ref. [40]), an additional update relation appears for the vector  $\bar{\mathbf{f}}_l$ . The above formulas are recursively used for  $l = 2, \dots, M - 1$ , using (5.23) for  $l = 1$ . The reflection and transmission coefficients may be computed when the cumulative matrix-vector for all  $M - 1$  interfaces is known

$$\begin{bmatrix} \mathbf{c}_M^+ \\ \mathbf{c}_1^- \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{S}}_{M-1}^{11} & \bar{\mathbf{S}}_{M-1}^{12} \\ \bar{\mathbf{S}}_{M-1}^{21} & \bar{\mathbf{S}}_{M-1}^{22} \end{bmatrix} \begin{bmatrix} \mathbf{c}_1^+ \\ \mathbf{c}_M^- \end{bmatrix} + \begin{bmatrix} \bar{\mathbf{f}}_{M-1}^1 \\ \bar{\mathbf{f}}_{M-1}^2 \end{bmatrix}. \quad (5.30)$$

In the contrast-field formulation of the FMM, the incoming field has been moved into a

source term. Therefore, the coefficients corresponding to incoming fields in the superstrate and substrate vanish,  $\mathbf{c}_1^+ = \mathbf{c}_M^- = \mathbf{0}$ , and we have

$$\begin{bmatrix} \mathbf{c}_M^+ \\ \mathbf{c}_1^- \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{f}}_{M-1}^1 \\ \bar{\mathbf{f}}_{M-1}^2 \end{bmatrix}. \quad (5.31)$$

Note that for the computation of the vector  $\bar{\mathbf{f}}_{M-1}$  in (5.29b) the blocks  $\bar{\mathbf{S}}_{l-1}^{11}$  and  $\bar{\mathbf{S}}_{l-1}^{21}$  are not required.

Very often, especially for aperiodic structures, we are also interested in the near field, i.e. the field inside and around the scatterer. For this purpose, also the intermediary coefficients in slices  $2, \dots, M-1$  need to be computed in a stable way. This requires the inverse of  $\mathbf{S}_l^{11}$  and consequently of  $\mathbf{X}_l$ , which causes instabilities due to growing exponentials. To avoid the direct inversion of  $\mathbf{X}_l$ , we separate the propagation matrices from the scattering matrix. Equation (5.19) is written as

$$\begin{bmatrix} \mathbf{c}_{l+1}^+ \\ \mathbf{c}_l^- \end{bmatrix} = \begin{bmatrix} \mathbf{S}_l'^{11} & \mathbf{S}_l'^{12} \\ \mathbf{S}_l'^{21} & \mathbf{S}_l'^{22} \end{bmatrix} \begin{bmatrix} \mathbf{X}_l \mathbf{c}_l^+ \\ \mathbf{X}_{l+1} \mathbf{c}_{l+1}^- \end{bmatrix} + \begin{bmatrix} \mathbf{f}_l^1 \\ \mathbf{f}_l^2 \end{bmatrix}. \quad (5.32)$$

In order to compute the intermediary coefficients a recursion from bottom to top is employed. From the first equation of (5.32) we can compute

$$\mathbf{X}_l \mathbf{c}_l^+ = (\mathbf{S}_l'^{11})^{-1} (\mathbf{c}_{l+1}^+ - \mathbf{S}_l'^{12} \mathbf{X}_{l+1} \mathbf{c}_{l+1}^- - \mathbf{f}_l^1), \quad (5.33)$$

which may be used in the second equation of (5.32) to compute  $\mathbf{c}_l^-$

$$\begin{aligned} \mathbf{c}_l^- &= \mathbf{S}_l'^{21} (\mathbf{S}_l'^{11})^{-1} \mathbf{c}_{l+1}^+ + (-\mathbf{S}_l'^{21} (\mathbf{S}_l'^{11})^{-1} \mathbf{S}_l'^{12} + \mathbf{S}_l'^{22}) \mathbf{X}_{l+1} \mathbf{c}_{l+1}^- \\ &\quad - \mathbf{S}_l'^{21} (\mathbf{S}_l'^{11})^{-1} \mathbf{f}_l^1 + \mathbf{f}_l^2. \end{aligned} \quad (5.34)$$

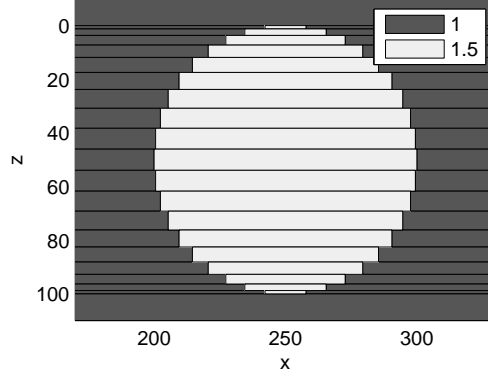
Finally, the  $\mathbf{c}_l^+$  is computed from the first equation of (5.24),

$$\mathbf{c}_l^+ = \bar{\mathbf{S}}_{l-1}^{11} \mathbf{c}_1^+ + \bar{\mathbf{S}}_{l-1}^{12} \mathbf{c}_l^- + \bar{\mathbf{f}}_{l-1}^1. \quad (5.35)$$

As before, due to the absence of an incoming field in the superstrate,  $\mathbf{c}_1^+ = \mathbf{0}$ , and (5.35) becomes

$$\mathbf{c}_l^+ = \bar{\mathbf{S}}_{l-1}^{12} \mathbf{c}_l^- + \bar{\mathbf{f}}_{l-1}^1. \quad (5.36)$$

Note that also in the computation of intermediary coefficients, the blocks  $\bar{\mathbf{S}}_{l-1}^{11}$  and  $\bar{\mathbf{S}}_{l-1}^{21}$  are not required. Thus, the computation of intermediary coefficients requires an additional sweep through the slices using recursive relations (5.34) and (5.36). The procedure



**Figure 5.3:** Sliced geometry of a cylinder. Different colors correspond to different refractive indices.

is formally described by the diagram

$$\mathbf{c}_{l+1} \xrightarrow{(\mathbf{s}_l, \mathbf{f}_l), (\bar{\mathbf{s}}_{l-1}, \bar{\mathbf{f}}_{l-1})} \mathbf{c}_l. \quad (5.37)$$

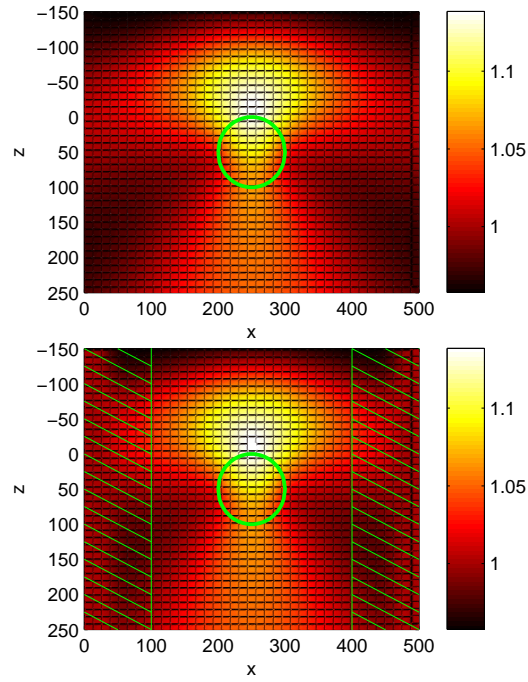
The recursion is initialized at  $l = M$ , with  $\mathbf{c}_M^- = \mathbf{0}$ ,  $\mathbf{c}_M^+ = \bar{\mathbf{f}}_{M-1}^1$ .

## 5.4 Numerical results

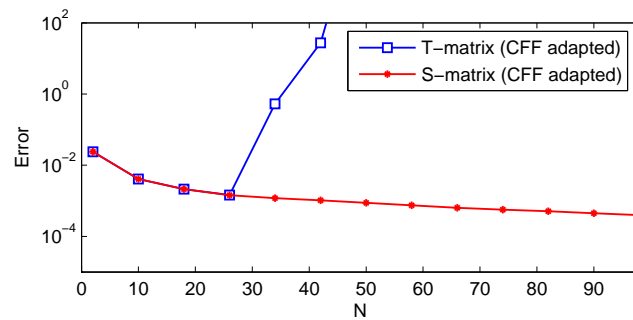
We consider the problem of scattering of a plane wave from a dielectric cylinder. An important argument for this choice is the fact that a semi-analytical solution can be found. The solution of the cylinder problem is obtained by writing the Maxwell equations in cylindrical coordinates, expanding the fields inside and outside the cylinder in terms of Bessel functions and finally matching the fields at the cylinder's interface [89].

In order to solve the same problem with AFMM-CFF, we first need to approximate the geometry by multiple rectangular slices. For a cylinder this may be done by imposing a fixed arc length between two adjacent slices. This will ensure an adaptive, slope-dependent staircasing of the profile. Figure 5.3 shows such a profile for a cylinder with radius  $\rho = 50$  nm obtained using  $M = 19$  slices. The incident plane wave has a wavelength  $\lambda = 628.3$  nm, travels downwards in the plane perpendicular to the cylinder and is TM-polarized. The semi-analytical solution and the corresponding AFMM-CFF solution of this problem are shown in Figure 5.4. The PMLs are placed in the stripes  $x \in [0, 100]$  and  $x \in [400, 500]$  and implemented using the coordinate transformation (4.54) with the parameters  $\beta_0 = 2$ ,  $p = 1$ . As expected, in the region between the PMLs the numerical solution is close to the reference solution.





**Figure 5.4:** Absolute values of the magnetic field: exact solution (top) and solution computed with AFMM-CFF (bottom). Hatched areas indicate PMLs.



**Figure 5.5:** Convergence of the AFMM-CFF with the non-homogeneous  $S$ -matrix and  $T$ -matrix approaches for a cylinder with radius  $\rho = 50$  nm approximated by  $M = 79$  slices.

	$\rho = 25$ nm	$\rho = 50$ nm	$\rho = 100$ nm
$N = 10$ , Non-hom. S-matrix	1.0261e+00	1.0600e+00	9.2628e-01
$N = 10$ , Non-hom. T-matrix	1.0261e+00	1.0600e+00	9.2628e-01
$N = 20$ , Non-hom. S-matrix	1.0264e+00	1.0602e+00	9.2614e-01
$N = 20$ , Non-hom. T-matrix	1.0264e+00	1.0602e+00	<i>1.2620e+00</i>
$N = 40$ , Non-hom. S-matrix	1.0265e+00	1.0602e+00	9.2629e-01
$N = 40$ , Non-hom. T-matrix	1.0265e+00	<i>1.4295e+00</i>	<i>6.5550e+19</i>
$N = 80$ , Non-hom. S-matrix	1.0266e+00	1.0602e+00	9.2636e-01
$N = 80$ , Non-hom. T-matrix	<i>1.0167e+00</i>	<i>3.6577e+19</i>	<i>6.1624e+61</i>
Reference	1.0266e+00	1.0602e+00	9.2637e-01

**Table 5.1:** Magnitudes of the magnetic field in a fixed point above the cylinder computed with the T-matrix and S-matrix algorithms adapted for CFF for increasing radius  $\rho$  and truncation order  $N$ .

We now turn our attention to the issue of stability. We use the cylinder problem to compare the performance of a non-homogeneous T-matrix algorithm and the non-homogeneous S-matrix algorithm presented in Section 5.3. Figure 5.5 shows the convergence of AFMM-CFF combined with these approaches for a cylinder with radius  $\rho = 50$  nm approximated by  $M = 79$  slices. The error is defined as

$$\mathcal{E} = \|h_y^{c,N}(x, z) - h_y^{c,\text{ref}}(x, z)\|_2 \text{ on } \Omega_s,$$

where  $h_y^{c,N}$  is the numerical solution obtained with AFMM-CFF for  $2N + 1$  harmonics,  $h_y^{c,\text{ref}}$  is the semi-analytical solution for the cylinder problem and  $\Omega_s = [200, 300] \times [0, 100]$  is the smallest rectangular domain enclosing the cylinder.

For a low truncation number  $N$ , both algorithms give similar results. At  $N > 30$  the T-matrix algorithm becomes unstable while the S-matrix algorithm yields accurate results for a larger number of harmonics. The truncation number for which the algorithm breaks down depends on the slice thickness, as demonstrated in Table 5.1. It lists the magnitudes of the magnetic field in a point (250, -100) computed with the two approaches. The radius of the cylinder  $\rho$  and the truncation order  $N$  are varying. From the matrix  $\mathbf{X}_l = e^{-k_0 \mathbf{Q}_l (h_{l+1} - h_l)}$  it is visible that increasing the radius of the cylinder while keeping a fixed number of slices (thus increasing slice thickness) or increasing the number of harmonics will lead to smaller entries in  $\mathbf{X}_l$  and respectively larger entries in its inverse, generating significant round-off errors. This explains the instabilities exhibited by the extended T-matrix approach for larger values of  $N$  and  $\rho$ . On the other hand, the non-homogeneous S-matrix algorithm is stable and gives correct results for all  $N$  and  $\rho$ . The offset of the computed solution with respect to the reference solution visible in Table 5.1 and the plateau on Figure 5.5 (error remains constant for  $N > 50$ ) are due to staircasing. This effect is well known for the FMM [67] and can be reduced by a normal vector field approach [78].



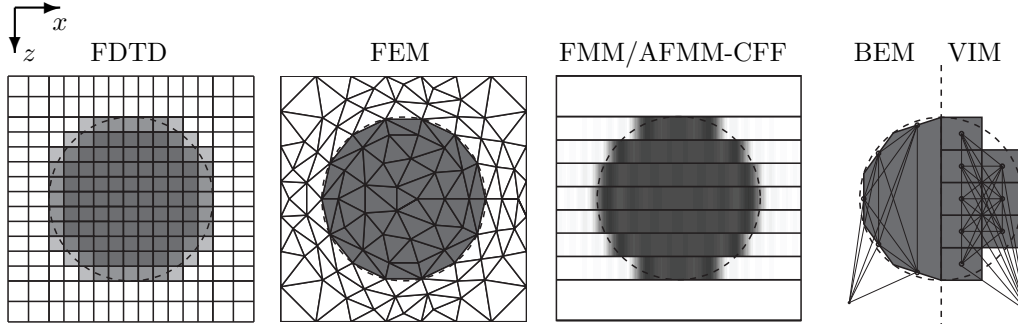
## Chapter 6

# Aperiodic Fourier modal method with alternative discretization

The aperiodic Fourier modal method in contrast-field formulation presented in the previous chapters uses spectral discretization (harmonics) in the finite periodic direction and spatial discretization (slices) in the other direction. In the light of the fact that the structures of interest have a large width-to-height ratio and that the two discretization approaches have different computational complexities, we propose exchanging the discretization directions. Moreover, if the scatterer has repeating patterns, exchanging the discretization directions facilitates the reuse of results of previous computations. Therefore the new method is suited for scattering from objects with a finite number of periods, such as gratings or memory arrays. The exchange of discretization directions requires a projection of the background field on the new basis introduced by the alternative discretization. Numerical experiments show that a considerable reduction of the computational costs (in terms of time and memory) can be obtained.

### 6.1 Introduction

Let us compare the discretization approach used by AFMM-CFF (and FMM) to the ones used by other popular methods such as FDTD, FEM, and integral equation methods. Figure 6.1 shows a two-dimensional cylinder problem discretized by these numerical methods. From the point of view of computational costs, the discretization process deter-



**Figure 6.1:** The rectangular grid of the FDTD method with sub-pixel smoothing to approximate the cylinder. The triangular grid of the FEM that conforms to the surface of the cylinder. The discrete layers of the FMM and AFMM-CFF, with the permittivity described by a few Fourier modes per layer. The source contributions of the surface elements for the BEM and the polarization densities in the cylinder for the VIM. Image reproduced from [28] with permission.

mines the number of degrees of freedom (DoFs) of the discrete problem. This number in turn determines the execution time and the memory requirements. For instance, a discretization of Maxwell equations (containing six unknowns) on a regular grid as in FDTD yields  $N_{\text{DOF}} = 6NM$  DoFs, where  $N$  and  $M$  represent the number of discretization points in the  $x$ - and  $z$ -direction respectively. If the system is subsequently solved by an iterative method, such as Gauss-Seidel, the computational cost is  $\mathcal{O}(N_{\text{DOF}}^2) = \mathcal{O}(N^2M^2)$ . Similar reasoning holds for the FEM and the integral equation methods, although the grid may be irregular in these cases. The key point here is that the computational cost of the methods mentioned above scales in the same way for all the discretization directions.

As can be seen from Figure 6.1, the FMM and AFMM-CFF make a distinction between the directions. In the vertical direction (the  $z$ -direction) the domain is divided into  $M$  layers in which the material properties are assumed to be  $z$ -independent. In each layer, the horizontal direction (the  $x$ -direction) is discretized by applying the Galerkin method with "shifted" Fourier harmonics as basis and test functions. This yields a system of ordinary differential equations of a size equal to the number of basis functions,  $N$ . The general solution (with unknown integration constants) of this system may be derived by numerically solving a matrix diagonalization problem. The diagonalization of dense matrices (based on **QR**-decomposition) has a computational complexity of  $\mathcal{O}(N^3)$  [19, Section 5.2]. Thus, for  $M$  layers the total computational cost is  $\mathcal{O}(N^3M)$ . Similarly the memory requirements can be estimated to  $\mathcal{O}(N^2M)$ .

It appears that the  $x$ -direction is "more expensive" both in terms of time and memory. For rectangular scatterers/domains that are much longer in the  $x$ -direction it is reasonable to choose an *alternative discretization*: make the longer direction "cheaper" by using spatial discretization into layers and apply spectral discretization in the shorter direction.

This exchange of directions turns out to be possible for the AFMM-CFF and not for the FMM since the latter requires different boundary conditions on the vertical and horizontal boundaries. For structures with repeating slices, the alternative discretization introduces slicing in the direction of repetition. We exploit this fact by proposing a new recursive algorithm which steps through the slices in a geometric progression. This further reduces the computational costs by replacing the linear scaling with the number of slices by a logarithmic one.

This chapter is structured as follows. In the next section we formulate a horizontal and an equivalent vertical problem whose solutions are to be computed with the AFMM-CFF using classical and alternative discretizations respectively. Section 6.3 discusses the computation of the background field for the the vertical problem. The final linear systems arising in the discretization of the vertical problem are derived in Section 6.4. In Section 6.5 we specifically consider structures with repeating patterns and present a fast recursive algorithm for solving the final linear system exploiting this property. A summary of the theoretical estimates of time and memory required by the AFMM-CFF is given in Section 6.6. The brief Section 6.7 contains a few remarks on the computation of the far field for the horizontal and vertical problems. Section 6.8 presents numerical evidence on the efficiency of the proposed approach. Practical speed-up factors are compared to the theoretical ones.

## 6.2 Two equivalent problems

We are looking for the solution of the time-harmonic Maxwell equations

$$\nabla \times \mathbf{e}(\mathbf{x}) = -k_0 \mathbf{h}(\mathbf{x}), \quad (6.1a)$$

$$\nabla \times \mathbf{h}(\mathbf{x}) = -k_0 \epsilon(x, z) \mathbf{e}(\mathbf{x}). \quad (6.1b)$$

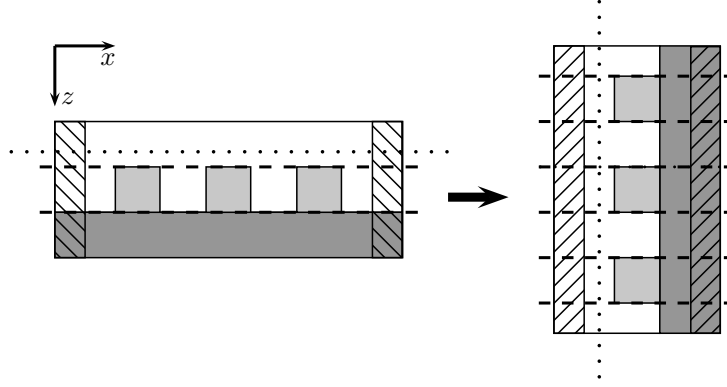
The incident field corresponds to a general conical incidence case and is given by

$$\mathbf{e}^{\text{inc}}(\mathbf{x}) = \mathbf{a}^e e^{-i\mathbf{k}^{\text{inc}} \cdot \mathbf{x}}. \quad (6.1c)$$

The numerical solution of (6.1) is to be computed on a domain  $\Omega = [0, \Lambda] \times \mathbb{R} \times \mathbb{R}$  using AFMM-CFF with classical and alternative discretization. Note that alternative discretization may be applied by a "turn" of the coordinate system

$$(x, y, z)^T \rightarrow (z, -y, x)^T, \quad (6.2)$$

such that formally the discretization directions remain unchanged, i.e. the  $x$ -direction is discretized with harmonics and the  $z$ -direction is discretized with slices in both approaches. Figure 6.2 demonstrates the turn for a finite grating with three rectangular



**Figure 6.2:** Alternative discretization by a turn of the coordinate system. Geometry of the horizontal problem  $\bar{\mathcal{P}}$  (left) and the vertical problem  $\bar{\bar{\mathcal{P}}}$  (right).

lines. Note that instead of three slices, we now have seven. From now on, we distinguish a *horizontal problem*  $\bar{\mathcal{P}}$  consisting of (6.1) with a geometry  $\epsilon(x, z) = \bar{\epsilon}(x, z)$  and an incoming field  $\mathbf{e}^{\text{inc}}(\mathbf{x}) = \bar{\mathbf{a}}e^{-i\bar{\mathbf{k}}^{\text{inc}} \cdot \mathbf{x}}$  and a *vertical problem*  $\bar{\bar{\mathcal{P}}}$  with  $\epsilon(x, z) = \bar{\bar{\epsilon}}(x, z)$  and  $\mathbf{e}^{\text{inc}}(\mathbf{x}) = \bar{\bar{\mathbf{a}}}e^{-i\bar{\bar{\mathbf{k}}}^{\text{inc}} \cdot \mathbf{x}}$ . We define a transformation matrix

$$\mathbf{L} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad (6.3)$$

which realizes the rotation of the problem. Then the vertical problem is related to the horizontal problem through

$$\bar{\bar{\epsilon}}(\mathbf{x}) = \bar{\epsilon}(\mathbf{L}\mathbf{x}), \quad (6.4a)$$

$$\bar{\bar{\mathbf{a}}} = \mathbf{L}\bar{\mathbf{a}}, \quad (6.4b)$$

$$\bar{\bar{\mathbf{k}}}^{\text{inc}} = \mathbf{L}\bar{\mathbf{k}}^{\text{inc}}, \quad (6.4c)$$

and the solutions of the two problems satisfy

$$\bar{\mathbf{e}}(\mathbf{x}) = \mathbf{L}\bar{\bar{\mathbf{e}}}(\mathbf{L}\mathbf{x}), \quad (6.5a)$$

$$\bar{\mathbf{h}}(\mathbf{x}) = \mathbf{L}\bar{\bar{\mathbf{h}}}(\mathbf{L}\mathbf{x}). \quad (6.5b)$$

In most applications the scatterer has a width which is much larger than its height which motivates the turn of coordinates. In the next section we describe application of the AFMM-CFF to the horizontal and vertical problems. Where necessary we discuss the differences and modifications required in order to solve the vertical problem  $\bar{\bar{\mathcal{P}}}$ .

### 6.3 Discretization of the background field

The background field is the solution to the problem of light propagation in a multilayer stack. In the horizontal problem (vertical problem), we assume a multilayer stack which extends infinitely in the  $xy$ -plane ( $zy$ -plane) and is invariant in the  $x$ - and  $y$ -directions ( $z$ - and  $y$ -directions). In each physical layer the field consists of an upward- and downward-traveling wave. Figure 6.3 depicts the background fields for the horizontal and vertical problems. The fundamental difference between the two cases lies in the fact that spatial discretization (slicing) is performed in a different physical direction. For the vertical background this new slicing introduces jumps in the permittivity functions per slice and as a result imposes a new set of basis functions per layer. We show in this section that the horizontal background field is represented exactly by a single basis function, while the vertical background field has to be approximated by a linear combination of the available basis functions.

The interfaces in the multilayer form a subset of the interfaces due to spatial discretization. We will therefore decompose the background field depending on discretization slices instead of physical layers,

$$\bar{\mathbf{e}}^b(x, y, z) = \bar{\mathbf{t}}_{s,l} e^{-k_0 \bar{q}_l (z-h_l)} e^{-i(\bar{k}_x^{\text{inc}} x + \bar{k}_y^{\text{inc}} y)} + \bar{\mathbf{r}}_{s,l} e^{k_0 \bar{q}_l (z-h_l)} e^{-i(\bar{k}_x^{\text{inc}} x + \bar{k}_y^{\text{inc}} y)}, \quad (6.6a)$$

$$\bar{\mathbf{h}}^b(x, y, z) = \bar{\mathbf{t}}_{u,l} e^{-k_0 \bar{q}_l (z-h_l)} e^{-i(\bar{k}_x^{\text{inc}} x + \bar{k}_y^{\text{inc}} y)} + \bar{\mathbf{r}}_{u,l} e^{k_0 \bar{q}_l (z-h_l)} e^{-i(\bar{k}_x^{\text{inc}} x + \bar{k}_y^{\text{inc}} y)}, \quad (6.6b)$$

for  $z \in [h_{l-1}, h_l)$ . Here

$$\bar{q}_l = i \sqrt{\epsilon_l^b - \left( \frac{k_x^{\text{inc}}}{k_0} \right)^2 - \left( \frac{k_y^{\text{inc}}}{k_0} \right)^2}, \quad (6.7)$$

and the amplitude vectors  $\mathbf{t}$  and  $\mathbf{r}$  can be determined from the interface conditions resulting from the Maxwell equations, see [92, 76].

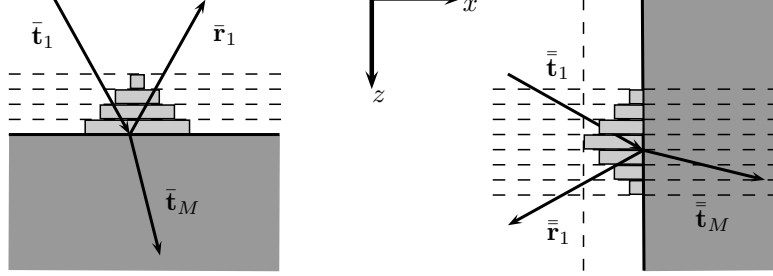
In the vertical problem the multilayer is invariant in the  $z$ - and  $y$ -directions. Application of the transformation (6.5) to (6.6) yields the vertical background field

$$\begin{aligned} \bar{\mathbf{e}}^b(x, y, z) &= \bar{\mathbf{t}}_{s,l} e^{-i\bar{k}_z^{\text{inc}} z} e^{-k_0 \bar{q}_l (x-h_l)} e^{-i\bar{k}_y^{\text{inc}} y} \\ &\quad + \bar{\mathbf{r}}_{s,l} e^{-i\bar{k}_z^{\text{inc}} z} e^{k_0 \bar{q}_l (x-h_l)} e^{-i\bar{k}_y^{\text{inc}} y}, \quad x \in [h_{l-1}, h_l), \end{aligned} \quad (6.8a)$$

$$\begin{aligned} \bar{\mathbf{h}}^b(x, y, z) &= \bar{\mathbf{t}}_{u,l} e^{-i\bar{k}_z^{\text{inc}} z} e^{-k_0 \bar{q}_l (x-h_l)} e^{-i\bar{k}_y^{\text{inc}} y} \\ &\quad + \bar{\mathbf{r}}_{u,l} e^{-i\bar{k}_z^{\text{inc}} z} e^{k_0 \bar{q}_l (x-h_l)} e^{-i\bar{k}_y^{\text{inc}} y}, \quad x \in [h_{l-1}, h_l), \end{aligned} \quad (6.8b)$$

where  $\bar{\mathbf{r}}_{u,l} = \mathbf{L} \bar{\mathbf{r}}_{u,l}$ ,  $\bar{\mathbf{r}}_{s,l} = \mathbf{L} \bar{\mathbf{r}}_{s,l}$ ,  $\bar{\mathbf{t}}_{u,l} = \mathbf{L} \bar{\mathbf{t}}_{u,l}$ ,  $\bar{\mathbf{t}}_{s,l} = \mathbf{L} \bar{\mathbf{t}}_{s,l}$ . We need to write the back-





**Figure 6.3:** The background field consisting of transmitted and reflected waves for the horizontal (left) and vertical (right) problem.

ground fields in a discretized form (4.57). It will be shown that the following less general form can be used for both settings (horizontal and vertical),

$$e_\alpha^b(x, y, z) = (\mathbf{s}_\alpha^b(z))^T \cdot \boldsymbol{\phi}(x, y) = (\mathbf{d}_{s,\alpha} s_\alpha^b(z))^T \cdot \boldsymbol{\phi}(x, y), \quad (6.9a)$$

$$h_\alpha^b(x, y, z) = (\mathbf{u}_\alpha^b(z))^T \cdot \boldsymbol{\phi}(x, y) = (\mathbf{d}_{u,\alpha} u_\alpha^b(z))^T \cdot \boldsymbol{\phi}(x, y). \quad (6.9b)$$

The horizontal background field can be exactly represented in this form. Since the  $(x, y)$ -dependent part of the horizontal background field (6.6) coincides with  $\phi_0(x, y)$ , we have  $\bar{\mathbf{d}}_{s,\alpha} = \bar{\mathbf{d}}_{u,\alpha} = \mathbf{e}_{N+1}$ , where  $\mathbf{e}_{N+1} \in \mathbb{R}^{2N+1}$  is defined as  $(\mathbf{e}_{N+1})_n = \delta_{n,N+1}$ , and

$$\bar{s}_\alpha^b(z) = \bar{t}_{s,\alpha,l} e^{-k_0 \bar{q}_l (z-h_l)} + \bar{r}_{s,\alpha,l} e^{k_0 \bar{q}_l (z-h_l)}, \quad z \in [h_{l-1}, h_l] \quad (6.10a)$$

$$\bar{u}_\alpha^b(z) = \bar{t}_{u,\alpha,l} e^{-k_0 \bar{q}_l (z-h_l)} + \bar{r}_{u,\alpha,l} e^{k_0 \bar{q}_l (z-h_l)}, \quad z \in [h_{l-1}, h_l] \quad (6.10b)$$

On the other hand, the vertical background field can only be represented in the form (4.57) or (6.9) through a projection. The basis function  $\phi_0(x, 0)$  gives the  $x$ -dependence for all plane waves constituting the background field in Figure 6.3 (left). None of the basis functions in the vertical problem give an exact representation of the background field in Figure 6.3 (right). In other words, in the vertical problem the basis used for the contrast field is not an eigenbasis for the background problem. We separate the  $x$ -dependent part of the vertical background field (6.8),

$$b_{s,\alpha}(x) = \bar{t}_{s,\alpha,l} e^{-k_0 \bar{q}_l (x-h_l)} + \bar{r}_{s,\alpha,l} e^{k_0 \bar{q}_l (x-h_l)}, \quad x \in [h_{l-1}, h_l], \quad (6.11a)$$

$$b_{u,\alpha}(x) = \bar{t}_{u,\alpha,l} e^{-k_0 \bar{q}_l (x-h_l)} + \bar{r}_{u,\alpha,l} e^{k_0 \bar{q}_l (x-h_l)}, \quad x \in [h_{l-1}, h_l], \quad (6.11b)$$

and project  $b_{s/u,\alpha}(x)$  on the available basis  $\phi_n(x, 0) = e^{-ik_{xn}x}$ ,

$$b_{s/u,\alpha}(x) = \sum_{n=-N}^N \hat{b}_{s/u,\alpha,n} e^{-ik_{xn}x}. \quad (6.12)$$

Multiplying both sides by  $e^{ik_{xn}x}$  and integrating over the interval  $[0, \Lambda]$ , yields

$$\hat{b}_{s/u,\alpha,n} = \frac{1}{\Lambda} \int_0^\Lambda b_{s/u}(x) e^{ik_{xn}x} dx = \frac{1}{\Lambda} \int_0^\Lambda (b_{s/u}(x) e^{ik_{x0}x}) e^{-in\frac{2\pi}{\Lambda}x} dx.$$

Thus the coefficients  $\hat{b}_{s/u,\alpha,n}$  may be computed with the help of a Fourier transform (FFT). We denote by  $\mathbf{b}_{s/u,\alpha}$  the vector formed from these coefficients. Now, the background field in the scatterer can be written in the available basis

$$\bar{e}_\alpha^{b,N}(x, y, z) = e^{-i\bar{k}_z^{\text{inc}}z} \sum_{n=-N}^N \hat{b}_{s,\alpha,n} \phi_n(x, y) = (\mathbf{b}_{s,\alpha} e^{-i\bar{k}_z^{\text{inc}}z})^T \boldsymbol{\phi}(x, y), \quad (6.13)$$

$$\bar{h}_\alpha^{b,N}(x, y, z) = e^{-i\bar{k}_z^{\text{inc}}z} \sum_{n=-N}^N \hat{b}_{u,\alpha,n} \phi_n(x, y) = (\mathbf{b}_{u,\alpha} e^{-i\bar{k}_z^{\text{inc}}z})^T \boldsymbol{\phi}(x, y). \quad (6.14)$$

The fields  $\bar{e}_\alpha^{b,N}$ ,  $\bar{h}_\alpha^{b,N}$  approximate the exact fields  $\bar{e}_\alpha^b$ ,  $\bar{h}_\alpha^b$  in (6.8) with  $2N + 1$  basis functions. Thus in representation (6.9) we have  $\bar{\mathbf{d}}_{s,\alpha} = \mathbf{b}_{s,\alpha}$ ,  $\bar{\mathbf{d}}_{u,\alpha} = \mathbf{b}_{u,\alpha}$  and

$$\bar{s}_\alpha^b(z) = \bar{u}_\alpha^b(z) = e^{-i\bar{k}_z^{\text{inc}}z}. \quad (6.15)$$

To achieve uniform notation in the horizontal and vertical settings, we denote

$$\bar{q}_l = i\bar{k}_z^{\text{inc}}. \quad (6.16)$$

From (6.10) and (6.15), we note that the following differential equation is satisfied for  $z \in [h_{l-1}, h_l]$

$$\frac{d}{dz^2} s_\alpha^b = q_l^2 s_\alpha^b, \quad (6.17a)$$

$$\frac{d}{dz^2} u_\alpha^b = q_l^2 u_\alpha^b. \quad (6.17b)$$

## 6.4 Recursive linear systems for the vertical problem

The discretization procedure in the AFMM-CFF applied to the vertical problem (alternative discretization) is similar to the one presented in Section 4.2 (for a horizontal problem). We list below the main differences:

- (i). The matrices related to the vertical background multilayer  $\mathbf{E}^b$  and  $\mathbf{P}^b$  become full (instead of diagonal matrices related to the horizontal background multilayer). The background permittivity in each slice has at least one jump discontinuity, such that its Fourier series requires more than one non-zero coefficient, as was the case for a

constant background permittivity per slice in the horizontal problem.

- (ii). For the horizontal problem the vector corresponding to the discretized background field (4.58) has a single non-zero entry

$$\mathbf{s}_{\alpha,l}^b(z) = \mathbf{d}_0 \bar{s}_{\alpha,l}^b(z) = \mathbf{d}_0 (a_{\alpha,l}^e e^{-k_0 \bar{q}_l(z-h_l)} + r_{\alpha,l}^e e^{k_0 \bar{q}_l(z-h_l)}), \quad (6.18a)$$

$$\mathbf{u}_{\alpha,l}^b(z) = \mathbf{d}_0 \bar{u}_{\alpha,l}^b(z) = \mathbf{d}_0 (a_{\alpha,l}^h e^{-k_0 \bar{q}_l(z-h_l)} + r_{\alpha,l}^h e^{k_0 \bar{q}_l(z-h_l)}). \quad (6.18b)$$

In the vertical problem, as a result of a projection, the "sparse" vector  $\mathbf{d}_0$  is replaced by a "dense vector"  $\mathbf{b}_{s/u,\alpha}$ .

$$\mathbf{s}_{\alpha,l}^b(z) = \mathbf{b}_{s,\alpha} \bar{s}_{\alpha,l}^b(z) = \mathbf{b}_{s,\alpha} e^{-k_0 \bar{q}_l(z-h_l)}, \quad (6.19a)$$

$$\mathbf{u}_{\alpha,l}^b(z) = \mathbf{b}_{u,\alpha} \bar{u}_{\alpha,l}^b(z) = \mathbf{b}_{u,\alpha} e^{-k_0 \bar{q}_l(z-h_l)}. \quad (6.19b)$$

The changes listed above are used in the derivations of the final linear systems for TE, TM and conical in Section 4.2 and the non-homogeneous terms  $\mathbf{g}_l$  in (4.74), (4.86) and (4.98) are accordingly modified. Note that the structure of the homogeneous part is not changed. A more detailed derivation for the conical case is given in [62]. We summarize these modifications below.

- **Planar incidence, TE-polarization**

The particular solution given by

$$\mathbf{s}_{y,\text{part},l}^c = \mathbf{p}_l e^{-k_0 \bar{q}_l(z-h_l)}, \quad (6.20)$$

where  $\mathbf{p}_l$  is computed by solving

$$(\mathbf{A}_l - \bar{q}_l^2 \mathbf{I}) \mathbf{p}_l = (\mathbf{E}_l - \mathbf{E}_l^b) \mathbf{b}_{s,y}. \quad (6.21)$$

The above equation is obtained by substitution of (6.20) in (4.66). After using the interface conditions (4.72) the non-homogeneous term in (4.74) becomes

$$\mathbf{g}_l(z) = \begin{bmatrix} \mathbf{p}_l \\ \mathbf{p}_l(-\bar{q}_l) \end{bmatrix} e^{-k_0 \bar{q}_l(z-h_l)}. \quad (6.22)$$

- **Planar incidence, TM-polarization**

The particular solution is given by

$$\mathbf{u}_{y,\text{part},l}^c = \mathbf{p}_l e^{-k_0 \bar{q}_l(z-h_l)}. \quad (6.23)$$

where  $\mathbf{p}_l$  is computed by solving

$$(\mathbf{P}_l^{-1} \mathbf{B}_l - \bar{q}_l^2 \mathbf{I}) \mathbf{p}_l = (-\mathbf{P}_l^{-1} \mathbf{B}_l - (\mathbf{P}_l^b)^{-1} \mathbf{B}_l^b) \mathbf{b}_{u,y}. \quad (6.24)$$

The non-homogeneous term in (4.86) becomes

$$\mathbf{g}_l(z) = \left[ \begin{array}{c} \mathbf{P}_l \\ (\mathbf{P}_l \mathbf{p}_l + (\mathbf{P}_l - \mathbf{P}_l^b) \mathbf{b}_{u,y}) (-\bar{q}_l) \end{array} \right] e^{-k_0 \bar{q}_l (z-h_l)}. \quad (6.25)$$

- **Conical incidence**

The particular solution is given by

$$\mathbf{s}_{x,\text{part},l}^c(z) = \mathbf{p}_{s,l} e^{-k_0 \bar{q}_l (z-h_l)}, \quad (6.26a)$$

$$\mathbf{u}_{x,\text{part},l}^c(z) = \mathbf{p}_{u,l} e^{-k_0 \bar{q}_l (z-h_l)}. \quad (6.26b)$$

Substituting this Ansatz in Equation (4.89) and using property (6.17), we obtain two linear systems which can be solved for  $\mathbf{p}_{s,l}$  and  $\mathbf{p}_{u,l}$

$$(\mathbf{C}_l - \bar{q}_l^2 \mathbf{I}) \mathbf{p}_{s,l} = -(\mathbf{B}_l \mathbf{P}_l^{-1} - \mathbf{B}_l^b (\mathbf{P}_l^b)^{-1}) \mathbf{b}_{s,x}, \quad (6.27a)$$

$$(\mathbf{D}_l - \bar{q}_l^2 \mathbf{I}) \mathbf{p}_{u,l} = (\mathbf{E}_l - \mathbf{E}_l^b) \mathbf{b}_{u,x}. \quad (6.27b)$$

The non-homogeneous term in (4.98) becomes

$$\mathbf{g}_l(z) = \left[ \begin{array}{c} \mathbf{P}_{s,l} \\ \mathbf{A}_l^{-1} (\mathbf{F} \mathbf{K}_x \mathbf{K}_y \mathbf{p}_{s,l} - k_0^{-1} \mathbf{p}_{u,l} \bar{q}_l + (\mathbf{E}_l - \mathbf{E}_l^b) \mathbf{b}_{s,y}) \\ \mathbf{P}_{u,l} \\ \mathbf{B}_l^{-1} (\mathbf{F} \mathbf{K}_x \mathbf{E}_l^{-1} \mathbf{K}_y \mathbf{p}_{u,l} - k_0^{-1} \mathbf{p}_{s,l} \bar{q}_l - i \mathbf{F} \mathbf{K}_x (\mathbf{I} - \mathbf{E}_l^b \mathbf{E}_l^{-1}) \mathbf{b}_{s,z}) \end{array} \right] e^{-k_0 \bar{q}_l (z-h_l)}, \quad (6.28)$$

with

$$\mathbf{b}_{s,z} = (\mathbf{E}_l^b)^{-1} (i \mathbf{F} \mathbf{K}_x \mathbf{b}_{u,y} - i \mathbf{K}_y \mathbf{b}_{u,x}). \quad (6.29)$$

The coupled linear systems (4.74), (4.86) and (4.98) with non-homogeneous terms (6.22), (6.25) and (6.28) can be solved in a stable manner by using the non-homogeneous S-matrix algorithm (see Chapter 5 and [63]) which is an adaptation of the classical (homogeneous) S-matrix algorithm to the contrast field formulation. However, it is shown in the next section that a dedicated algorithm can be developed which has lower time and memory requirements.

## 6.5 Non-homogeneous S-matrix algorithm for repeating slices

In this section we develop a dedicated non-homogeneous S-matrix algorithm for the vertical problem. The repeating patterns in the direction of slicing suggest that this property can be exploited in order to achieve even higher efficiency in terms of computational costs.

### 6.5.1 Redheffer notation

As explained in Chapter 5 the interface relations (4.74), (4.86) and (4.98) are unstable in the sense that its direct use for sequential elimination of field amplitudes leads to large round-off errors [40, 63]. To avoid instability, we use an S-matrix representation that maps the incoming waves on an interface to outgoing waves from that interface. In this chapter we will make use of a Redheffer star product for matrices [71] which simplifies the formulation of the proposed algorithm and allows grouping (or merging) of interface into one equivalent interface described by a single S-matrix. It then becomes natural to denote the S-matrix connecting layers  $a$  and  $b$  ( $a < b$ ,  $a, b \in \mathbb{N}$ ) by  $\mathbf{S}_{a,b}$  and the corresponding source vector by  $\mathbf{f}_{a,b}$ . We refer to this notation as *Redheffer notation*. In this section it replaces the standard notation used in Chapter 5. Particularly, the S-matrix and T-matrix corresponding to a single interface between layers  $l$  and  $l+1$  are denoted by  $\mathbf{S}_{l,l+1}$  and  $\mathbf{T}_{l,l+1}$  (Redheffer notation) instead of  $\mathbf{S}_l$  and  $\mathbf{T}_l$  respectively (standard notation). For convenience the S-matrix relations for a single interface presented in Section 5.3 are listed here again using the Redheffer notation. The S-matrix relation associated with layers  $l$  and  $l+1$  is (see Figure 6.4)

$$\begin{bmatrix} \mathbf{c}_{l+1}^+ \\ \mathbf{c}_l^- \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{S}_{l,l+1}^{11} & \mathbf{S}_{l,l+1}^{12} \\ \mathbf{S}_{l,l+1}^{21} & \mathbf{S}_{l,l+1}^{22} \end{bmatrix}}_{\mathbf{S}_{l,l+1}} \begin{bmatrix} \mathbf{c}_l^+ \\ \mathbf{c}_{l+1}^- \end{bmatrix} + \underbrace{\begin{bmatrix} \mathbf{f}_{l,l+1}^1 \\ \mathbf{f}_{l,l+1}^2 \end{bmatrix}}_{\mathbf{f}_{l,l+1}}.$$

The pair  $(\mathbf{S}_{l,l+1}, \mathbf{f}_{l,l+1})$  describes the scattering and source properties of the interface. It can be determined from the T-matrix  $\mathbf{T}_{l,l+1}$  and the vector  $\mathbf{g}'_{l,l+1}$  using the relations

$$\mathbf{S}_{l,l+1}^{11} = (\mathbf{T}'_{l,l+1}{}^{11} - \mathbf{T}'_{l,l+1}{}^{12} (\mathbf{T}'_{l,l+1}{}^{22})^{-1} \mathbf{T}'_{l,l+1}{}^{21}) \mathbf{X}_l, \quad (6.30a)$$

$$\mathbf{S}_{l,l+1}^{12} = \mathbf{T}'_{l,l+1}{}^{12} (\mathbf{T}'_{l,l+1}{}^{22})^{-1} \mathbf{X}_{l+1}, \quad (6.30b)$$

$$\mathbf{S}_{l,l+1}^{21} = -(\mathbf{T}'_{l,l+1}{}^{22})^{-1} \mathbf{T}'_{l,l+1}{}^{21} \mathbf{X}_l, \quad (6.30c)$$

$$\mathbf{S}_{l,l+1}^{22} = (\mathbf{T}'_{l,l+1}{}^{22})^{-1} \mathbf{X}_{l+1}, \quad (6.30d)$$

and

$$\mathbf{f}_{l,l+1}^1 = \mathbf{g}'_{l,l+1}{}^1 - \mathbf{T}'_{l,l+1}{}^{12} (\mathbf{T}'_{l,l+1}{}^{22})^{-1} \mathbf{g}'_{l,l+1}{}^2, \quad (6.30e)$$

$$\mathbf{f}_{l,l+1}^2 = -(\mathbf{T}'_{l,l+1}{}^{22})^{-1} \mathbf{g}'_{l,l+1}{}^2. \quad (6.30f)$$

Here

$$\begin{bmatrix} \mathbf{T}'_{l,l+1}{}^{11} & \mathbf{T}'_{l,l+1}{}^{12} \\ \mathbf{T}'_{l,l+1}{}^{21} & \mathbf{T}'_{l,l+1}{}^{22} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \mathbf{W}_{l+1}^{-1} & \mathbf{V}_{l+1}^{-1} \\ \mathbf{W}_{l+1}^{-1} & -\mathbf{V}_{l+1}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{W}_l & \mathbf{W}_l \\ \mathbf{V}_l & -\mathbf{V}_l \end{bmatrix}, \quad (6.31a)$$

$$\begin{bmatrix} \mathbf{g}'_{l,l+1}{}^1 \\ \mathbf{g}'_{l,l+1}{}^2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \mathbf{W}_{l+1}^{-1} & \mathbf{V}_{l+1}^{-1} \\ \mathbf{W}_{l+1}^{-1} & -\mathbf{V}_{l+1}^{-1} \end{bmatrix} (\mathbf{g}_l(h_{l+1}) - \mathbf{g}_{l+1}(h_{l+1})). \quad (6.31b)$$

Matrices  $\mathbf{W}_l$  and  $\mathbf{V}_l$  are computed by eigenvalue decomposition, which is a relatively expensive computation. The time and memory requirements are estimated by

$$\mathcal{T}_{\text{diag}} = \mathcal{O}(M'N^3), \quad (6.32a)$$

$$\mathcal{M}_{\text{diag}} = \mathcal{O}(M'N^2). \quad (6.32b)$$

Here  $M'$  is the number of interfaces per period.

## 6.5.2 Non-homogeneous Redheffer star product

We now consider the problem of replacing two neighboring interfaces by one equivalent interface, depicted in Figure 6.5. Let the pair  $(\mathbf{S}_{a,b}, \mathbf{f}_{a,b})$  describe the interface(s) joining layers  $a$  and  $b$ , and the pair  $(\mathbf{S}_{b,c}, \mathbf{f}_{b,c})$  describe the interface(s) joining layers  $b$  and  $c$ , ( $a < b < c$ ,  $a, b, c \in \mathbb{N}$ ).

$$\begin{bmatrix} \mathbf{c}_b^+ \\ \mathbf{c}_a^- \end{bmatrix} = \begin{bmatrix} \mathbf{S}_{a,b}^{11} & \mathbf{S}_{a,b}^{12} \\ \mathbf{S}_{a,b}^{21} & \mathbf{S}_{a,b}^{22} \end{bmatrix} \begin{bmatrix} \mathbf{c}_a^+ \\ \mathbf{c}_b^- \end{bmatrix} + \begin{bmatrix} \mathbf{f}_{a,b}^1 \\ \mathbf{f}_{a,b}^2 \end{bmatrix}, \quad (6.33a)$$

$$\begin{bmatrix} \mathbf{c}_c^+ \\ \mathbf{c}_b^- \end{bmatrix} = \begin{bmatrix} \mathbf{S}_{b,c}^{11} & \mathbf{S}_{b,c}^{12} \\ \mathbf{S}_{b,c}^{21} & \mathbf{S}_{b,c}^{22} \end{bmatrix} \begin{bmatrix} \mathbf{c}_b^+ \\ \mathbf{c}_c^- \end{bmatrix} + \begin{bmatrix} \mathbf{f}_{b,c}^1 \\ \mathbf{f}_{b,c}^2 \end{bmatrix}. \quad (6.33b)$$

We would like to find the pair  $(\mathbf{S}_{a,c}, \mathbf{f}_{a,c})$  which realizes the mapping for layers  $a$  and  $c$

$$\begin{bmatrix} \mathbf{c}_c^+ \\ \mathbf{c}_a^- \end{bmatrix} = \begin{bmatrix} \mathbf{S}_{a,c}^{11} & \mathbf{S}_{a,c}^{12} \\ \mathbf{S}_{a,c}^{21} & \mathbf{S}_{a,c}^{22} \end{bmatrix} \begin{bmatrix} \mathbf{c}_a^+ \\ \mathbf{c}_c^- \end{bmatrix} + \begin{bmatrix} \mathbf{f}_{a,c}^1 \\ \mathbf{f}_{a,c}^2 \end{bmatrix}. \quad (6.34)$$

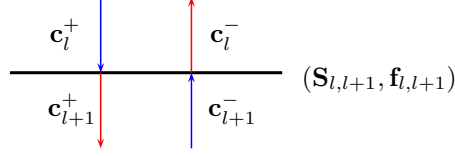


Figure 6.4: Interface description.

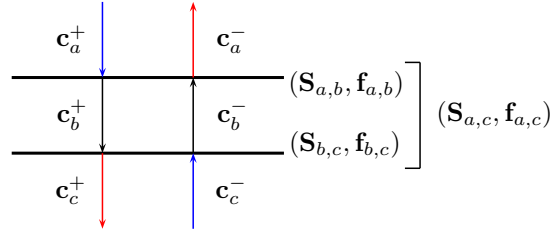


Figure 6.5: The problem of interface merging.

Similarly to the derivation in Section 5.3 it can be shown that  $(\mathbf{S}_{a,c}, \mathbf{f}_{a,c})$  is given by

$$\mathbf{S}_{a,c} = \begin{bmatrix} \mathbf{S}_{b,c}^{11}(\mathbf{I} - \mathbf{S}_{a,b}^{12}\mathbf{S}_{b,c}^{21})^{-1}\mathbf{S}_{a,b}^{11} & \mathbf{S}_{b,c}^{12} + \mathbf{S}_{b,c}^{11}\mathbf{S}_{a,b}^{12}(\mathbf{I} - \mathbf{S}_{b,c}^{21}\mathbf{S}_{a,b}^{12})^{-1}\mathbf{S}_{b,c}^{22} \\ \mathbf{S}_{a,b}^{21} + \mathbf{S}_{a,b}^{22}\mathbf{S}_{b,c}^{21}(\mathbf{I} - \mathbf{S}_{a,b}^{12}\mathbf{S}_{b,c}^{21})^{-1}\mathbf{S}_{a,b}^{11} & \mathbf{S}_{a,b}^{22}(\mathbf{I} - \mathbf{S}_{b,c}^{21}\mathbf{S}_{a,b}^{12})^{-1}\mathbf{S}_{b,c}^{22} \end{bmatrix}, \quad (6.35)$$

$$\mathbf{f}_{a,c} = \begin{bmatrix} \mathbf{S}_{b,c}^{11}(\mathbf{I} - \mathbf{S}_{a,b}^{12}\mathbf{S}_{b,c}^{21})^{-1}(\mathbf{S}_{a,b}^{12}\mathbf{f}_{b,c}^2 + \mathbf{f}_{a,b}^1) + \mathbf{f}_{b,c}^1 \\ \mathbf{S}_{a,b}^{22}\mathbf{S}_{b,c}^{21}(\mathbf{I} - \mathbf{S}_{a,b}^{12}\mathbf{S}_{b,c}^{21})^{-1}(\mathbf{S}_{a,b}^{12}\mathbf{f}_{b,c}^2 + \mathbf{f}_{a,b}^1) + \mathbf{S}_{a,b}^{22}\mathbf{f}_{b,c}^2 + \mathbf{f}_{a,b}^2 \end{bmatrix}. \quad (6.36)$$

To simplify further exposition, based on the classical *Redheffer star product* for matrices [71], we define a *non-homogeneous Redheffer star product* for matrix-vector pairs

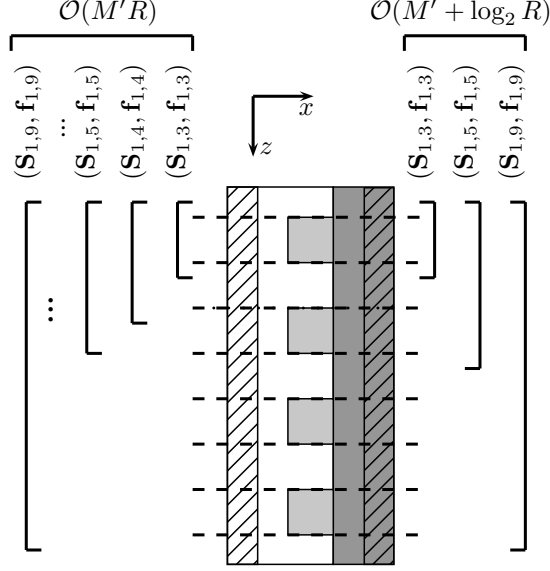
$$(\mathbf{S}_{a,b}, \mathbf{f}_{a,b}) * (\mathbf{S}_{b,c}, \mathbf{f}_{b,c}) = (\mathbf{S}_{a,c}, \mathbf{f}_{a,c}). \quad (6.37)$$

It can be proven that this new product, just like the regular matrix product and the Redheffer matrix product, is associative but not commutative. We will exploit the associativity property in order to group the interfaces in a convenient way.

### 6.5.3 Global stack description

We can repeat the process of merging interfaces until the whole stack of interfaces is described by one equivalent interface. The non-homogeneous S-matrix algorithm (derived in Section 5.3) uses the iteration

$$(\mathbf{S}_{1,l+1}, \mathbf{f}_{1,l+1}) = (\mathbf{S}_{1,l}, \mathbf{f}_{1,l}) * (\mathbf{S}_{l,l+1}, \mathbf{f}_{l,l+1}). \quad (6.38)$$



**Figure 6.6:** A schematic representation of the nonhomogeneous  $S$ -matrix algorithm applied to a grating with  $R = 4$  periods and  $M' = 2$  interfaces per period: classical linear recursion (left side) and fast exponential recursion (right side).

The global matrix-vector pair for  $M - 1$  interfaces ( $M$  layers) is given by

$$(\mathbf{S}_{1,M}, \mathbf{f}_{1,M}) = [\dots [ [ (\mathbf{S}_{1,2}, \mathbf{f}_{1,2}) * (\mathbf{S}_{2,3}, \mathbf{f}_{2,3}) ] * (\mathbf{S}_{3,4}, \mathbf{f}_{3,4}) ] * \dots * (\mathbf{S}_{M-1,M}, \mathbf{f}_{M-1,M})]. \quad (6.39)$$

The left side of Figure 6.6 gives a visual representation of the merging process.

We consider a structure with  $R$  repeating periods for simplicity assumed to be a power of 2,  $R = 2^K$  (the algorithm can be extended to an arbitrary number of periods), where each period has  $M'$  interfaces in the discretization ( $M - 1 = RM'$ ).

The cost of computing (6.39) scales linearly with the number of periods,

$$\mathcal{T}_S = \mathcal{O}(RM'N^3), \quad (6.40a)$$

$$\mathcal{M}_S = \mathcal{O}(RM'N^2). \quad (6.40b)$$

Due to the associativity property of the non-homogeneous Redheffer product, we can regroup the multiplication operations. This allows us to exploit the local periodicity and



reduce the computational costs down to

$$\mathcal{T}_S = \mathcal{O}((M' + \log_2(R))N^3), \quad (6.41a)$$

$$\mathcal{M}_S = \mathcal{O}((M' + \log_2(R))N^2). \quad (6.41b)$$

We observe that once we have computed the matrix-vector corresponding to the stack  $\{1, \dots, M' + 1\}$ , the matrix-vector corresponding to the stack  $\{M' + 1, \dots, 2M' + 1\}$  can be computed with a minimal effort

$$(\mathbf{S}_{M'+1, 2M'+1}, \mathbf{f}_{M'+1, 2M'+1}) = (\mathbf{S}_{1, M'+1}, \nu \mathbf{f}_{1, M'+1}), \quad (6.42)$$

where  $\nu = e^{-k_0 q (h_{M'+1} - h_1)}$  is a phase factor which appears due to the scalar  $z$ -dependence of  $\mathbf{f}$  (see (6.22), (6.25), and (6.28)). Thus, we can directly compute the matrix-vector corresponding to the stack  $\{1, \dots, 2M' + 1\}$

$$(\mathbf{S}_{1, 2M'+1}, \mathbf{f}_{1, 2M'+1}) = (\mathbf{S}_{1, M'+1}, \mathbf{f}_{1, M'+1}) * (\mathbf{S}_{1, M'+1}, \nu \mathbf{f}_{1, M'+1}). \quad (6.43)$$

The iteration process is defined by

$$\begin{aligned} (\mathbf{S}_{1, 2^{k+1}M'+1}, \mathbf{f}_{1, 2^{k+1}M'+1}) &= (\mathbf{S}_{1, 2^k M'+1}, \mathbf{f}_{1, 2^k M'+1}) \\ &* (\mathbf{S}_{1, 2^k M'+1}, \nu^{2^k} \mathbf{f}_{1, 2^k M'+1}), \end{aligned} \quad (6.44)$$

for  $k = 0, \dots, K - 1$ .

Once the matrix-vector pair for the whole stack is computed, the coefficients in the upper and lower layers are given by

$$\begin{pmatrix} \mathbf{c}_M^+ \\ \mathbf{c}_1^- \end{pmatrix} = \begin{pmatrix} \mathbf{f}_{1, M}^1 \\ \mathbf{f}_{1, M}^2 \end{pmatrix}. \quad (6.45)$$

#### 6.5.4 Intermediary coefficients

Relation (6.45) allows the computation of the coefficients in the first and last layers of the stack. In order to compute the coefficients in the intermediary layers, we start with the pair associated with the whole stack and recursively halve the stack. One iteration of this process is explained by again considering the problem in Figure 6.5. This time we are given the pairs  $(\mathbf{S}_{a,b}, \mathbf{f}_{a,b})$ ,  $(\mathbf{S}_{b,c}, \mathbf{f}_{b,c})$  and the amplitudes  $\mathbf{c}_a^+$ ,  $\mathbf{c}_c^-$ , and we need to determine the intermediary coefficients  $\mathbf{c}_b^+$  and  $\mathbf{c}_b^-$ . Substitution of the second equation of (6.33b) in the first equation of (6.33a) yields

$$(\mathbf{I} - \mathbf{S}_{a,b}^{12} \mathbf{S}_{b,c}^{21}) \mathbf{c}_b^+ = \mathbf{S}_{a,b}^{11} \mathbf{c}_a^+ + \mathbf{S}_{a,b}^{12} \mathbf{S}_{b,c}^{22} \mathbf{c}_c^- + \mathbf{S}_{a,b}^{12} \mathbf{f}_{b,c}^2 + \mathbf{f}_{a,b}^1. \quad (6.46)$$

Substitution of the first equation of (6.33a) in the second equation of (6.33b) yields

$$(\mathbf{I} - \mathbf{S}_{b,c}^{21} \mathbf{S}_{a,b}^{12}) \mathbf{c}_b^- = \mathbf{S}_{b,c}^{21} \mathbf{S}_{a,b}^{11} \mathbf{c}_a^+ + \mathbf{S}_{b,c}^{22} \mathbf{c}_c^- + \mathbf{S}_{b,c}^{21} \mathbf{f}_{a,b}^1 + \mathbf{f}_{b,c}^2. \quad (6.47)$$

We can now write the amplitudes in layer  $b$  in terms of the amplitudes of the incoming modes in layer  $a$  and  $c$ ,

$$\begin{pmatrix} \mathbf{c}_b^+ \\ \mathbf{c}_b^- \end{pmatrix} = \begin{pmatrix} (\mathbf{I} - \mathbf{S}_{a,b}^{12} \mathbf{S}_{b,c}^{21})^{-1} \mathbf{S}_{a,b}^{11} & \mathbf{S}_{a,b}^{12} (\mathbf{I} - \mathbf{S}_{b,c}^{21} \mathbf{S}_{a,b}^{12})^{-1} \mathbf{S}_{b,c}^{22} \\ \mathbf{S}_{b,c}^{21} (\mathbf{I} - \mathbf{S}_{a,b}^{12} \mathbf{S}_{b,c}^{21})^{-1} \mathbf{S}_{a,b}^{11} & (\mathbf{I} - \mathbf{S}_{b,c}^{21} \mathbf{S}_{a,b}^{12})^{-1} \mathbf{S}_{b,c}^{22} \end{pmatrix} \begin{pmatrix} \mathbf{c}_a^+ \\ \mathbf{c}_c^- \end{pmatrix} + \begin{pmatrix} \mathbf{S}_{a,b}^{12} \mathbf{f}_{b,c}^2 + \mathbf{f}_{a,b}^1 \\ \mathbf{S}_{b,c}^{21} \mathbf{f}_{a,b}^1 + \mathbf{f}_{b,c}^2 \end{pmatrix}. \quad (6.48)$$

The update relation (6.48) is applied recursively (by halving the stack) until the coefficients in all layers are determined. This requires  $RM' - 1$  evaluations of (6.48) which, if computed in the right order, involves only matrix-vector products. Therefore, the computational cost of computing and storing the intermediary coefficients is given by

$$\mathcal{T}_c = \mathcal{O}(RM'N^2), \quad (6.49a)$$

$$\mathcal{M}_c = \mathcal{O}(RM'N). \quad (6.49b)$$

## 6.6 Summary of computational costs

From the point of view of computational cost the AFMM-CFF with classical and alternative discretizations consists of three consecutive tasks:

- (i). Matrix diagonalization (or eigenvalue decomposition);
- (ii). Calculation of the global S-matrix;
- (iii). Calculation of the intermediary coefficients.

The computational costs (in terms of time and memory) for each of the tasks is given respectively by (6.32), (6.41) and (6.49). By combining these estimates we obtain the total computational costs,

$$\mathcal{T} = ((c_{\text{diag}}^{\mathcal{T}} + c_S^{\mathcal{T}})M' + c_S^{\mathcal{T}} \log_2(R))N^3 + c_c^{\mathcal{T}} RM'N^2, \quad (6.50a)$$

$$\mathcal{M} = ((c_{\text{diag}}^{\mathcal{M}} + c_S^{\mathcal{M}})M' + c_S^{\mathcal{M}} \log_2(R))N^2 + c_c^{\mathcal{M}} RM'N. \quad (6.50b)$$

We specify that  $R$  in the above expressions gives the number of repeating patterns in the vertical direction. Thus, for horizontally repeating patterns we always have  $R = 1$  and  $M'$  represents the total number of interfaces. With this convention, the estimate

(6.50) is general and applicable for the vertical as well as the horizontal problem. For a reasonably large  $N$ , the second term in the above expressions (being one order lower) can be neglected. We will also assume (supported by evidence from practical experiments) that  $c_{\text{diag}}^{\mathcal{T}} = c_S^{\mathcal{T}} = c^{\mathcal{T}}$ . Then estimate (6.50) reduces to

$$\mathcal{T} \approx (2c^{\mathcal{T}} M' + c^{\mathcal{T}} \log_2(R)) N^3, \quad (6.51a)$$

$$\mathcal{M} \approx ((c_{\text{diag}}^{\mathcal{M}} + c_S^{\mathcal{M}}) M' + c_S^{\mathcal{M}} \log_2(R)) N^2. \quad (6.51b)$$

## 6.7 Far-field recovery

In practical applications it is often needed to compute the field far above the scatterer. The far field can be computed by a Green's function approach [44] or by a Fourier transformation of the field on a line above the scatterer, denoted by a dotted line in Figure 6.2). The second approach [20, Section 3.10.2] is faster since the integrals can be computed numerically with fast Fourier transform (FFT) routines. From Figure 6.2 it is visible that the field on the dotted line is not available in the PMLs and beyond for the horizontal problem. Therefore, the Fourier transformation approach can only be used in the vertical problem.

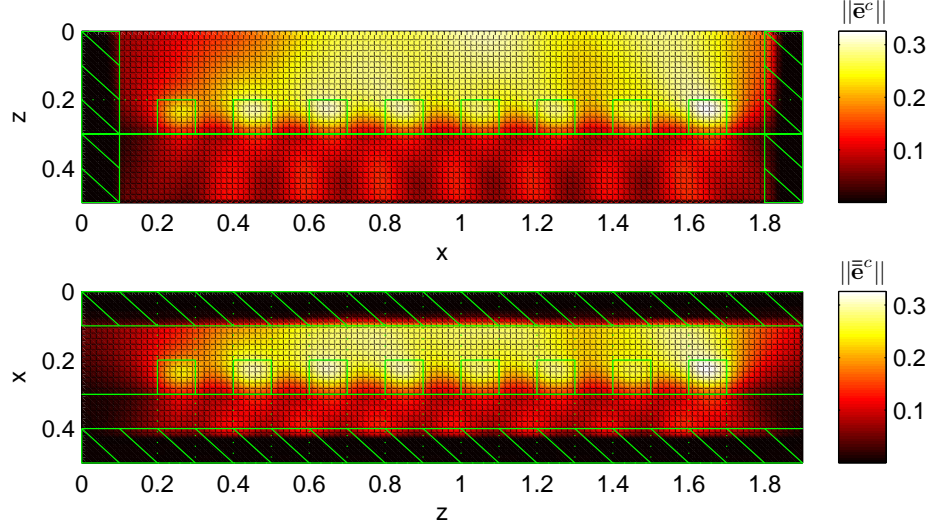
## 6.8 Numerical results

We consider the problem of scattering from finite gratings. These are periodic structures with a finite number of periods. The problem is inspired from real-life applications in lithography, where gratings are printed on the wafer to be later used for quality control. By measuring the light scattered from a grating, its shape can be determined and the quality of the lithographic process can be assessed.

We first consider a small grating, infinitely long in the  $y$ -direction. The geometry of the problem is plotted in Figure 6.7. It consists of  $R = 8$  rectangular lines made of resist (with a refractive  $n_{\text{resist}} = 1.5$ ), referred to as scatterer, supported by a silicon substrate ( $n_{\text{silicon}} = 4.28 - 0.05i$ ). The material above the substrate and the grating lines is air ( $n_{\text{air}} = 1.0$ ). The refractive indices  $n$  of the material are related to the electric permittivity through

$$\epsilon(x, z) = n^2(x, z). \quad (6.52)$$

We obtain the numerical solution by using the classical and alternative discretization approaches for AFMM-CFF or, as explained in Section 6.2, by solving a horizontal problem



**Figure 6.7:** Near field at  $y = 0$  for a binary grating with 8 lines computed with classical (top) and alternative discretization (bottom).

and a vertical problem. For the horizontal problem, the incident plane wave is given by (6.1c) with

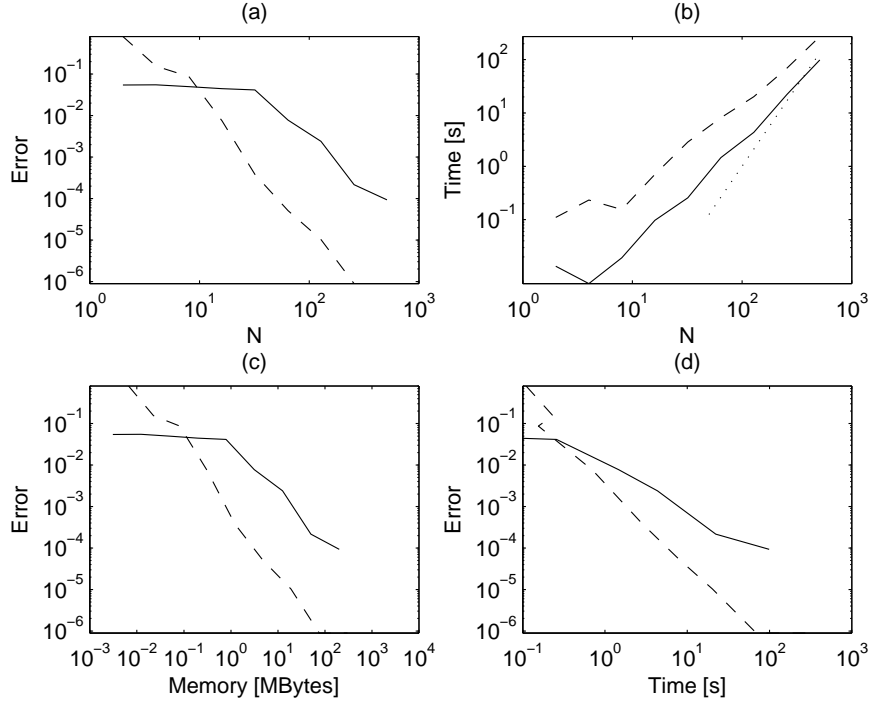
$$\begin{aligned}\bar{\mathbf{a}} &= (0, 1, 0)^T, \\ \bar{\mathbf{k}}^{\text{inc}} &= 10(\sqrt{2}/2, 0, \sqrt{2}/2)^T,\end{aligned}$$

and the PMLs are placed in the regions  $x \in [0, 0.1] \cup [1.8, 1.9]$ . The distance units are  $\mu\text{m}$ . For the vertical problem, the amplitude and the wave vectors of the incident plane wave are given by

$$\begin{aligned}\bar{\bar{\mathbf{a}}} &= (0, -1, 0)^T, \\ \bar{\bar{\mathbf{k}}}^{\text{inc}} &= 10(\sqrt{2}/2, 0, \sqrt{2}/2)^T,\end{aligned}$$

and the PMLs are placed in the regions  $x \in [0, 0.1] \cup [0.4, 0.5]$ . It follows from the wavevectors  $\bar{\mathbf{k}}^{\text{inc}}$  that in both cases the incident plane wave is TE-polarized and has a wavelength  $\lambda = 2\pi/|\bar{\mathbf{k}}^{\text{inc}}| \approx 0.6283$ . Note that the spatial discretization requires  $\bar{M} = 3$  slices in the horizontal problem and  $\bar{\bar{M}} = \bar{R}\bar{M}' + 1 = 17$  slices in the vertical problem.

Figure 6.7 shows the solutions of the two problems. The solution of the vertical problem is plotted in rotated coordinates  $(z, -y, x)$  to make easier the comparison with the horizontal problem. Clearly, for both problems the fields decay to zero in the PMLs. The solution in the PML is not physically relevant and the placement of the PMLs is



**Figure 6.8:** Comparison of classical and alternative discretization approaches for a grating with 32 rectangular lines on resist substrate: (a) Convergence, (b) Complexity, (c) Error vs. Memory, (d) Error vs. Time. Solid line (—) corresponds to the horizontal problem (classical discretization), dashed line (- -) corresponds to the vertical problem (alternative discretization), dotted line ( $\cdots$ ) indicates theoretical cubic complexity,  $\mathcal{T} = \mathcal{O}(N^3)$ .

different in the two problems. Therefore, the horizontal problem has a relevant solution in the stripe  $(x, y, z) \in \bar{\Omega}_r = [0.1, 1.8] \times \mathbb{R} \times \mathbb{R}$ , while the vertical problem has a relevant solution on  $(z, -y, x) \in \bar{\bar{\Omega}}_r = \mathbb{R} \times \mathbb{R} \times [0.1, 0.4]$ . In the light of the above discussion, the fields should be compared only on a domain where both are relevant, that is on

$$\Omega_c = \bar{\Omega}_r \cup \bar{\bar{\Omega}}_r = [0.1, 1.8] \times \mathbb{R} \times [0.1, 0.4]. \quad (6.53)$$

Thus we expect that

$$\bar{\mathbf{e}}^{c,N}(\mathbf{x}) \approx \mathbf{L} \bar{\mathbf{e}}^{c,N}(\mathbf{L}\mathbf{x}) \text{ for } \mathbf{x} = (x, y, z) \in \Omega_c, \quad (6.54)$$

which is confirmed by a visual inspection of the obtained solutions in  $\Omega_c$ . Note that  $N$  in (6.54) may be different for the horizontal and vertical problems.

In order to clearly assess the advantages of alternative discretization over classical discretization, we need to consider convergence and complexity of the two approaches. Figure 6.8 (a) shows the convergence of the solution for a vertical and horizontal setup. The error definitions are given by

$$\bar{\mathcal{E}}^N = \|\bar{\mathbf{e}}^{c,N}(\mathbf{x}) - \bar{\mathbf{e}}^{c,\text{ref}}(\mathbf{x})\|, \quad (6.55)$$

$$\bar{\bar{\mathcal{E}}}^N = \|\mathbf{L}\bar{\mathbf{e}}^{c,N}(\mathbf{L}\mathbf{x}) - \bar{\mathbf{e}}^{c,\text{ref}}(\mathbf{x})\|, \quad (6.56)$$

where the norm for a vector field  $\mathbf{v}$  is defined as

$$\|\mathbf{v}\|^2 = \sum_i \sum_j \sum_\alpha |v_\alpha(x_i, 0, z_j)|^2, \quad (6.57)$$

for  $\alpha \in \{x, y, z\}$  and  $x_i$  and  $z_j$  sample the smallest rectangular domain enclosing the scatterer. For scatterers invariant in the  $y$ -direction, the dependence of the field on  $y$  is known, and only the field at  $y = 0$  may be considered. In absence of an exact solution for this type of problems, we choose as reference the numerical solution for the (trusted) horizontal setup computed with a large truncation number  $N$

$$\mathbf{e}^{c,\text{ref}} = \bar{\mathbf{e}}^{c,1048}.$$

The decay length in the PML is set to  $1/\sigma_0 = 1/6$ . The computations are performed using MATLAB on a Intel Core 2 Quad CPU Q6600 4x2.40GHz with 2GB of RAM.

Figure 6.8 (a) shows that for small truncation numbers such as  $N < 10$ , the error for the vertical problem is larger. The reason for this behavior lies in the representation of the vertical background field in the available basis. While in the horizontal problem a single basis function is required for an exact representation, in the vertical problem the background field  $\mathbf{e}^b, \mathbf{h}^b$  can only be approximated. Clearly, for small  $N$  the accuracy of the approximation is unsatisfactory. The slow decrease of the error for both approaches is explained by an insufficient resolution of the basis functions, so that the exponential decay of the solution in the PMLs cannot be accurately resolved. For  $N > 10$ , the error for the vertical problem  $\bar{\mathcal{E}}$  becomes much smaller than for the horizontal one  $\bar{\bar{\mathcal{E}}}$ . Because the width of the computational domain for  $\bar{\mathcal{P}}$  is smaller than the width of the computational domain for  $\bar{\bar{\mathcal{P}}}$ , less basis functions are needed for the former to achieve the same accuracy. From the plot (Figure 6.8 (a)) we conclude that the same number of harmonics determine an error for the vertical problem which is two orders of magnitude smaller than for the horizontal problem. However this plot is not sufficient to judge the performance of the approaches. Indeed, for a fixed number of harmonics, more work needs to be done for the vertical problem than for the horizontal due to a larger number of slices.

We turn to Figure 6.8 (b) which shows the execution time as a function of  $N$  for the two problems. Since  $\bar{M}' = 2, \bar{R} = 1$  and  $\bar{\bar{M}}' = 2, \bar{\bar{R}} = 32$ , using (6.51) we can estimate the

ratio of computation times for a large (but fixed)  $N$ :

$$\frac{(2c^{\mathcal{T}}\bar{M}' + c^{\mathcal{T}}\log_2(\bar{R}))\bar{N}^3}{(2c^{\mathcal{T}}\bar{M}' + c^{\mathcal{T}}\log_2(\bar{R}))\bar{N}^3} = \frac{4+0}{4+5} = \frac{1}{2.25}.$$

Thus solving the vertical problem is estimated to take 2.25 times longer than the horizontal problem. The distance between the two lines on the plot for  $N$  approaching  $10^3$  comes close to this value. We also note that although asymptotically we expect a cubic complexity  $\mathcal{T} = \mathcal{O}(N^3)$  indicated by the dotted line in the plot, it is only approached for  $N > 10^2$ . This means that the real speed-up factor will be lower than the one predicted theoretically.

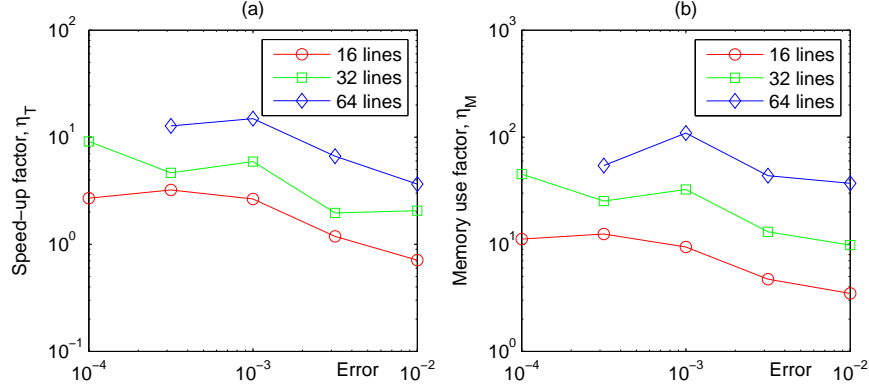
A technical detail is in order. The values of the matrix  $\mathbf{X}_l$  defined in (4.73a) can become extremely small (for very thick layers or large number of harmonics) which leads to the issue of *underflow* and *subnormal numbers* [54, Chapter 7]. Arithmetic operations with subnormal numbers take up to ten times longer than with regular numbers in floating-point representation, which would yield a supercubic behavior in Figure 6.8 (b). To avoid the issue we flush these numbers to zero. The accuracy of the results is not influenced and the computational time is reduced.

We combine data from Figure 6.8 (a) and (b) in order to obtain the direct dependence of error on time displayed in Figure 6.8 (d). We conclude that for errors larger than  $10^{-2}$  the vertical approach is slower. It only becomes faster for smaller errors. For instance, an error of  $10^{-3}$  can be achieved approximately 8 times faster by solving  $\bar{\bar{\mathcal{P}}}$  instead of  $\bar{\mathcal{P}}$ .

For large computations with AFMM-CFF, the memory requirements limit the number of harmonics to be used. For this reason we also consider the memory required by the horizontal and vertical approaches to achieve a certain error. The size of a matrix or vector element in a double precision floating-point representation is 8 bytes. Considering that we have four matrices per layer, the estimated memory space required for a computation is given by (6.50b) with  $c_{\text{diag}}^{\mathcal{M}} = c_S^{\mathcal{M}} = 36$  and  $c_c^{\mathcal{M}} = 16$ . Thus, the information from the convergence plot, Figure 6.8 (a), can be used to determine the error-vs-memory behavior displayed in Figure 6.8 (c).

We expect the gain in computational costs to change depending on the width of the structures. For this reason we now consider three geometries with the same material properties but having a different number of periods. Using the functions  $\bar{\mathcal{T}}(\mathcal{E})$  and  $\bar{\bar{\mathcal{T}}}(\mathcal{E})$ , we define the *speed-up factor* for a given error

$$\eta_{\mathcal{T}}(\mathcal{E}) = \frac{\bar{\mathcal{T}}(\mathcal{E})}{\bar{\bar{\mathcal{T}}}(\mathcal{E})}. \quad (6.58)$$



**Figure 6.9:** Speed-up and memory use factors for  $R \in \{16, 32, 64\}$ ,  $n_s = 1.5$ .

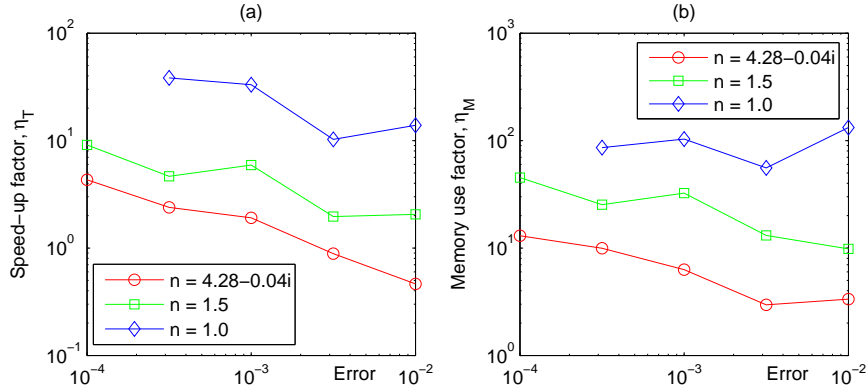
In analogy with the speed-up factor we also define the *memory use factor*

$$\eta_{\mathcal{M}}(\mathcal{E}) = \frac{\bar{\mathcal{M}}(\mathcal{E})}{\underline{\mathcal{M}}(\mathcal{E})}. \quad (6.59)$$

A plot of the speed-up and memory use factor for different values of  $R$  ( $R \in \{16, 32, 64\}$ ) is shown in Figure 6.9. We observe the general trend of increasing factors for decreasing errors. For  $\mathcal{E} = 10^{-3}$  the speed-up and memory use factors are in the range  $3 \dots 15$  and  $10 \dots 110$  respectively. Unfortunately it is difficult to simulate larger gratings ( $R \in \{128, 256, \dots\}$ ) with classical discretization (horizontal problem) due to the prohibitively large (in terms of memory) number of harmonics. This is also the reason why the error of  $10^{-4}$  has not been reached for  $R = 64$  (it is however easily reached in the vertical problem, see Figure 6.8). We can extrapolate the data presented in the plot in order to predict an approximate speed-up of  $10^2$  and a memory use factor of  $10^3$  at  $\mathcal{E} = 10^{-3}$  for a grating having 256 lines.

The convergence of the solution depends strongly on its smoothness, which in turn is determined by the presence and the size of the jump discontinuities in the discretized permittivity function  $\epsilon_l(x)$ . Clearly, these functions are different for the vertical and horizontal problems. The jump discontinuities are material interfaces which can be due to the scatterer and due to the background multilayer. Independent of the discretization direction, the discontinuities introduced by the scatterer have the same effect on the jumps in  $\epsilon_l(x)$ . The effect of the background multilayer however is different depending on the discretization direction. For the horizontal problem the jumps in  $\bar{\epsilon}^b(z)$  are not visible in  $\bar{\epsilon}_l(x)$  since they coincide with the interfaces of discretization slices. On the other hand, in the vertical problem the jump discontinuities in  $\bar{\epsilon}^b(x)$  are present in  $\bar{\epsilon}_l(x)$  for all layers. This might slow down the convergence when solving the vertical problem. To





**Figure 6.10:** Speed-up and memory use factors for  $R = 32$ ,  $n_s \in \{1.0, 1.5, 4.28 - 0.04i\}$ .

quantify the effect, we consider a grating with  $R = 32$  lines with three different materials in the substrate: air ( $n_s = 1.0$ ), resist ( $n_s = 1.5$ ), silicon ( $n_s = 4.28 - 0.04i$ ). A plot of the speed-up and memory use factor for these cases is shown in Figure 6.10. In the case of air substrate (which is less relevant for metrology applications) the background permittivity function  $\epsilon^b$  is continuous, which determines a faster convergence of the vertical problem's solution. This results in large speed-up and memory use factors (up to 40 and 100 respectively). On the other hand, for silicon substrate we have a large jump discontinuity in  $\epsilon^b$  at the interface between air and silicon and much smaller gains (2 and 8 for speed and memory respectively).

We can get a theoretical estimate of the attainable speed-up using (6.58) and (6.51)

$$\eta_{\mathcal{T}}(\mathcal{E} \rightarrow 0) = \frac{2\bar{M}' + \log_2(\bar{R})}{2\bar{M}' + \log_2(\bar{R})} \left( \frac{\bar{N}}{\bar{N}} \right)^3. \quad (6.60)$$

The argument  $\mathcal{E} \rightarrow 0$  is used in order to stress the fact that this estimate is accurate for large  $N$  (that is small  $\mathcal{E}$ ). We consider the case of a resist grating with 32 lines on air substrate studied in Figure 6.10. Because the superstrate and substrate are of the same material we have similar smoothness properties of the solution and permittivity in both directions. Then, we may assume that in order to achieve the same accuracy, we need to have equal resolution of the discretization. Thus, we require that the number of harmonics per unit length is constant for both discretizations. This yields the estimate

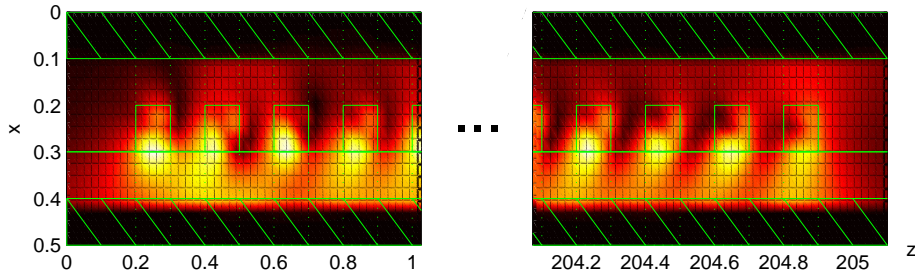
$$\bar{N}/\bar{N} = \bar{\Lambda}/\bar{\Lambda} = 67/5 = 13.4. \quad (6.61)$$

We also have  $\bar{M}' = 2$ ,  $\bar{R} = 1$  and  $\bar{M}' = 2$ ,  $\bar{R} = 32$ . Then, the speed-up factor estimated with (6.60) is close to 1000. We stress that this is a theoretical estimation for very large  $N$  or, equivalently, for extremely small  $\mathcal{E}$ . The estimate is useful as an upper bound for

a potentially attainable speed-up. In practice, the pure cubic complexity in  $N$  is hardly reached. For the considered problem, the error of  $10^{-3}$  is reached for  $N = 10^1 \dots 10^2$ . In this range a quadratic complexity can be assumed, as confirmed by Figure 6.8 (b). In this case our estimated speed-up factor becomes

$$\eta_{\mathcal{T}}(\mathcal{E} \approx 10^{-3}) = \frac{2\bar{M}' + \log_2(\bar{R})}{2\bar{M}' + \log_2(\bar{R})} \left( \frac{\bar{N}}{\bar{N}} \right)^2 = 79.8, \quad (6.62)$$

which is reasonably close to the measured  $\eta_{\mathcal{T}}$  for  $n_s = 1.0$  shown in Figure 6.10. It is important to stress that convergence depends on the smoothness of the permittivity. For problems which have different smoothness properties in different directions, the requirement on equal resolution (harmonics per unit length) in order to achieve equal errors (used in (6.61)) is not applicable.



**Figure 6.11:** Plot of  $\|\bar{\mathbf{e}}^c\|$  for an ultra-large grating with 1024 lines. Only small parts of the field at the edges of the grating are shown.

Although the most noticeable improvement resulting from the alternative discretization is speed-up, the memory saving might become crucial when solving large problems. Thus, with the new approach we may easily simulate scattering from structures with repeating patterns having a width of the order of hundreds of wavelengths. This is a difficult task for the AFMM-CFF with classical discretization, as well as for other numerical methods such as FDTD and FEM. Figure 6.11 shows a small part of the computed field for a grating with 1024 lines (more than 300 wavelengths).



## Chapter 7

# Conclusions and suggestions for future work

We have presented an extension of the FMM which enables simulation of scattering from two-dimensional finite structures (invariant in the third dimension) illuminated by plane waves at arbitrary angles of incidence. A detailed derivation has been provided for the case of planar incidence and TE-polarization which was subsequently generalized to planar TM and conical incidence as well as to arbitrary shapes of the scatterer requiring more than three slices in the spatial discretization. The Galerkin approach was used for the discretization of the first-order time-harmonic Maxwell equations. The formulation in terms of a contrast field resembles the scattered field formulations used in the curvilinear coordinate method [18] as well as in FEM and FDTD [83]. However, the reformulation for the FMM is less straightforward, since it requires solutions which can be written in analytical form. As a result of the reformulation, the FMM had to be adapted to solve non-homogeneous equations of second order. Besides the application demonstrated in this thesis, this allows modeling of internal sources inside the domain. We mention that a method of solving non-homogeneous equations of first order with the FMM has been previously demonstrated by Bai and Turunen [2]. The convergence study in Chapter 3 has shown that for the aperiodic model problem the AFMM-CFF needs less harmonics than the supercell FMM. In view of the fact that the number of operations performed by the eigenvalue solver (which is computationally the most demanding step in the method) scales cubically with the number of harmonics, this results in a considerable reduction of computational time. In our example, for an accuracy of  $10^{-3}$ , a speed-up by a factor of  $8^3$  can be reached.

We have shown that in the FMM discretization leads to homogeneous linear systems, while in the AFMM-CFF the source term causes the resulting linear systems to be-

come non-homogeneous. The existing S-matrix algorithm has been adapted for use with non-homogeneous recursive linear systems. For this purpose, a set of recursive update relations has been derived which give a stable algorithm for the computation of reflection and transmission coefficients. Moreover, a backward recursion may be employed to determine the intermediary coefficients if needed. The stability of the developed method has been qualified by comparison against a semi-analytical result obtained for the problem of scattering of a plane wave from a dielectric cylinder.

Finally, a technique for speeding up computations with the AFMM-CFF has been presented. The speed-up is achieved by exchanging the directions of spatial and spectral discretizations. The cheaper spatial discretization is used in the longer direction, while the more expensive spectral discretization is applied in the shorter direction. Moreover, periodicity in the slicing direction is exploited by using the associativity of the non-homogeneous Redheffer star product, which further decreases the computational costs. The exchange of discretization directions is implemented by a rotation of the coordinate system. While the rotated scatterer can be treated automatically by the method, the rotated background multilayer introduces a fundamental difference: the background field cannot be exactly represented in the available basis and a projection has to be used. This step introduces an additional approximation error, especially when few basis functions are used. For this reason, speed-up is achieved when smaller errors are required (or equivalently more basis functions are employed). Besides the imposed accuracy, the speed-up and memory use factors obtained with alternative discretization are dependent on the number of repeating patterns as well as the size of jump discontinuities in the background multilayer. The numerical experiments confirm that the alternative discretization shows larger improvements for geometries with smaller jumps at the layer interfaces and a larger number of periods.

The research described in this thesis has highlighted several open questions related to the standard FMM as well as opened possibilities for further extension of the newly developed AFMM-CFF. Concerning theoretical results on the convergence of the FMM we mention the proof of Li for the TE case [3]. No such proof exists for a more general case. The formulation of the method using the Galerkin approach exposed in Chapter 4 gives access to mathematical tools which could be useful for a rigorous convergence proof for the FMM. Ref. [24] discusses this and other issues related to the mathematical foundation of the FMM.

A more practical issue is the computational cost of the method. Currently the FMM and AFMM-CFF compute the general solution ( $e^{-\sqrt{\mathbf{A}}z} + e^{\sqrt{\mathbf{A}}z}$ ) of a second-order ODE by matrix diagonalizations. It might be possible to reduce these costs by considering other approaches for numerical evaluation of matrix exponentials. Ref. [49] describes nineteen different algorithms to compute the matrix exponential ranging from the Taylor series approximation to multistep ODE solvers. At the cost of lower accuracy, iterative approaches might prove to be faster than the explicit diagonalization approach.

The choice of basis functions used in the discretization directly depends on the angle of the incident plane wave and the period  $\Lambda$ . In the AFMM-CFF,  $\Lambda$  represents the period of the artificially periodized finite structure. This value can be increased (at the cost of larger number of harmonics and computation times) without affecting the solution. This gives extra freedom in choosing the basis functions and can be exploited to impose a basis which more accurately represents several incident plane-waves or the background field in the AFMM-CFF with alternative discretization.

In this thesis we have considered two-dimensional scatterers (finite in one direction). The presented approach can be extended to three-dimensional scatterers supported on multilayers invariant in the  $x$ - and  $y$ -directions. This extension would allow accurate modeling of scatterers which are finite in two directions. The double-periodic FMM is a good starting point for the extension. In this case two pairs of PMLs have to be used in the two finite directions. Alternative discretization (presented in Chapter 6) could also be applied for three-dimensional geometries. In this case instead of the classical approach (spectral discretization of  $x$ - and  $y$ - directions and spatial discretization of  $z$ ) we can choose either  $x$  or  $y$  to be spatially discretized.

Finally, we mention that the FMM and AFMM-CFF with minimal changes can be applied to problems of acoustic scattering from respectively infinitely periodic and finite structures.



# Appendix A

## Vector calculus identities

$$\nabla \cdot (\nabla \times \mathbf{v}) = 0 \tag{A.1}$$

$$\nabla \cdot (f\mathbf{v}) = \mathbf{v} \cdot \nabla f + f\nabla \cdot \mathbf{v} \tag{A.2}$$

$$\nabla \times (f\mathbf{v}) = f\nabla \times \mathbf{v} - \mathbf{v} \times \nabla f \tag{A.3}$$

$$\nabla \cdot (\mathbf{v}_1 \times \mathbf{v}_2) = (\nabla \times \mathbf{v}_1) \cdot \mathbf{v}_2 - \mathbf{v}_1 \cdot (\nabla \times \mathbf{v}_2) \tag{A.4}$$





# Appendix B

## Derivations

### B.1 Derivation of the convolution term

$$\begin{aligned} & \sum_{n=-\infty}^{\infty} \hat{\epsilon}_{n,l} e^{in\frac{2\pi}{\Lambda}x} \sum_{m=-\infty}^{\infty} s_m(z) e^{-ik_{xn}x} = \\ & \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \hat{\epsilon}_n s_m(z) e^{-i(k_x^{\text{inc}} + (n+m)\frac{2\pi}{d}x)} = \quad [p = n + m] \\ & \sum_{p=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \hat{\epsilon}_{p-m} s_m(z) e^{-i(k_x^{\text{inc}} + p\frac{2\pi}{d}x)} = \\ & \sum_{p=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \hat{\epsilon}_{p-m} s_m(z) e^{-ik_{xp}x}. \end{aligned}$$

### B.2 Discretization using the Galerkin approach

We demonstrate the Galerkin approach on the Equation (4.4f). Discretization of the other Maxwell equations follows by analogy.

$$\frac{\partial}{\partial x} h_{y,l} - \frac{\partial}{\partial y} h_{x,l} = -k_0 \epsilon_l(x) e_{z,l}. \quad (\text{B.1})$$

The basis functions are given by

$$\phi_n(x, y) = e^{-i(k_{xn}x + k_y y)}, \quad (\text{B.2})$$

where

$$k_{xn} = k_x^{\text{inc}} - n \frac{2\pi}{\Lambda}, \quad k_y = k_y^{\text{inc}}, \quad \text{for } n = -N \dots + N.$$

In each slice  $l$  the fields are expanded as

$$e_{\alpha,l}(x, y, z) = \sum_{n=-N}^N s_{\alpha,l,n}(z) \phi_n(x, y) = (\mathbf{s}_{\alpha,l}(z))^T \cdot \boldsymbol{\phi}(x, y), \quad (\text{B.3a})$$

$$h_{\alpha,l}(x, y, z) = \sum_{n=-N}^N u_{\alpha,l,n}(z) \phi_n(x, y) = (\mathbf{u}_{\alpha,l}(z))^T \cdot \boldsymbol{\phi}(x, y). \quad (\text{B.3b})$$

Substitute (B.3) in (B.1)

$$\begin{aligned} & \sum_{n=-N}^N u_{y,l,n}(z) (-ik_{xn}) e^{-i(k_{xn}x + k_y y)} - \sum_{n=-N}^N u_{x,l,n}(z) (-ik_y) e^{-i(k_{xn}x + k_y y)} \\ &= -k_0 \epsilon_l(x) \sum_{n=-N}^N s_{z,l,n}(z) e^{-i(k_{xn}x + k_y y)}. \end{aligned} \quad (\text{B.4})$$

We define the inner product on  $[0, \Lambda)$

$$\langle f, g \rangle = \int_0^\Lambda f(x) \bar{g}(x) dx, \quad (\text{B.5})$$

where  $\bar{g}$  denotes the complex conjugate of  $g$ . Using this inner product, Equation (B.4) is projected on each of the basis functions  $e^{-i(k_{xm}x + k_y y)}$ .

$$\begin{aligned} & \sum_{n=-N}^N u_{y,l,n}(z) (-ik_{xn}) \int_0^\Lambda e^{-i(k_{xn}x + k_y y)} e^{i(k_{xm}x + k_y y)} dx \\ & - \sum_{n=-N}^N u_{x,l,n}(z) (-ik_y) \int_0^\Lambda e^{-i(k_{xn}x + k_y y)} e^{i(k_{xm}x + k_y y)} dx \\ &= -k_0 \sum_{n=-N}^N s_{z,l,n}(z) \int_0^\Lambda \epsilon_l(x) e^{-i(k_{xn}x + k_y y)} e^{i(k_{xm}x + k_y y)} dx, \quad m = -N \dots N. \end{aligned} \quad (\text{B.6})$$

Note that the basis functions are orthogonal

$$\begin{aligned} \int_0^\Lambda e^{-i(k_{xn}x+k_yy)} e^{i(k_{xm}x+k_yy)} dx &= \int_0^\Lambda e^{-i(k_{xn}-k_{xm})x} dx \\ &= \int_0^\Lambda e^{i(n-m)\frac{2\pi}{\Lambda}x} dx = \Lambda\delta_{mn}, \quad m = -N \dots N. \end{aligned}$$

Equation (B.6) reduces to

$$u_{y,l,m}(z)(-ik_{xm}) - u_{x,l,m}(z)(-ik_y) \tag{B.7}$$

$$= -k_0 \sum_{n=-N}^N s_{z,l,n}(z) \int_0^\Lambda \epsilon_l(x) e^{i(n-m)\frac{2\pi}{\Lambda}x} dx, \quad m = -N \dots N. \tag{B.8}$$

We recognize the integral on the right-hand side to be the  $(n-m)$ -th Fourier coefficient  $\hat{\epsilon}_{l,n-m}$  of the permittivity function  $\epsilon_l$ .

$$u_{y,l,m}(z)(-ik_{xm}) - u_{x,l,m}(z)(-ik_y) \tag{B.9}$$

$$= -k_0 \sum_{n=-N}^N s_{z,l,n}(z) \hat{\epsilon}_{l,n-m}, \quad m = -N \dots N. \tag{B.10}$$

Finally we can write the above system of equations in matrix form

$$-i\mathbf{K}_x \mathbf{u}_{y,l}(z) + i\mathbf{K}_y \mathbf{u}_{x,l}(z) = -\mathbf{E}_l \mathbf{s}_{z,l}(z), \tag{B.11}$$

where

$$(\mathbf{K}_x)_{mn} = (k_{xn}/k_0)\delta_{mn}, \tag{B.12a}$$

$$(\mathbf{K}_y)_{mn} = (k_y/k_0)\delta_{mn}, \tag{B.12b}$$

$$(\mathbf{E}_l)_{mn} = \hat{\epsilon}_{l,n-m}. \tag{B.12c}$$



# Bibliography

- [1] Uri M. Ascher, Robert M. M. Mattheij, and Robert D. Russell. *Numerical Solution of Boundary Value Problems for Ordinary Differential Equations*. Cambridge University Press, 1995.
- [2] Benfeng Bai and Jari 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(5):1105–1112, May 2007.
- [3] Gang Bao, Lawrence Cowsar, and Wen Masters, editors. *Mathematical Modeling in Optical Science (Frontiers in Applied Mathematics)*. Society for Industrial Mathematics, January 2001.
- [4] Rudolf Beck and Ralf Hiptmair. Multilevel solution of the time-harmonic Maxwell's equations based on edge elements. *International Journal for Numerical Methods in Engineering*, 45(7):901–920, 1999.
- [5] J. Berenger. A perfectly matched layer for the absorption of electromagnetic waves. *Journal of Computational Physics*, 114(2):185–200, October 1994.
- [6] J. P. Berenger. Perfectly matched layer for the FDTD solution of wave-structure interaction problems. *IEEE Transactions on Antennas and Propagation*, 44(1):110–117, January 1996.
- [7] M. Botha. Solving the volume integral equations of electromagnetic scattering. *Journal of Computational Physics*, 218(1):141–158, October 2006.
- [8] J. J. Bowman, T. B. A. Senior, and P. L. E. Uslenghi. *Electromagnetic and acoustic scattering by simple shapes (Revised edition)*. New York, Hemisphere Publishing Corp., 1987.
- [9] Nicolas Chateau and Jean-Paul Hugonin. Algorithm for the rigorous coupled-wave analysis of grating diffraction. *J. Opt. Soc. Am. A*, 11(4):1321–1331, April 1994.
- [10] W. C. Chew, J. M. Jin, and E. Michielssen. Complex Coordinate Stretching as a Generalized Absorbing Boundary Condition. *Microwave and Optical Technology Letters*, 15:363–369, 1997.

- 
- [11] W. C. Chew and W. H. Weedon. A 3D perfectly matched medium from modified Maxwell's equations with stretched coordinates. *Microwave and Optical Technology Letters*, 7:599–604, September 1994.
- [12] F. Collino and P. Monk. Optimizing the perfectly matched layer. *Computer Methods in Applied Mechanics and Engineering*, 164(1-2):157–171, October 1998.
- [13] Francis Collino and Peter Monk. The Perfectly Matched Layer in Curvilinear Coordinates. *SIAM J. Sci. Comput.*, 19:2061–2090, November 1998.
- [14] R. Courant, K. Friedrichs, and H. Lewy. Über die partiellen Differenzgleichungen der mathematischen Physik. *Mathematische Annalen*, 100(1):32–74, December 1928.
- [15] Timothy A. Davis. *Direct Methods for Sparse Linear Systems (Fundamentals of Algorithms)*. Society for Industrial and Applied Mathematics, illustrated edition, September 2006.
- [16] Alexei Deinega and Ilya Valuev. Subpixel smoothing for conductive and dispersive media in the finite-difference time-domain method. *Opt. Lett.*, 32(23):3429–3431, December 2007.
- [17] Lewis F. DeSandre and J. Merle Elson. Extinction-theorem analysis of diffraction anomalies in overcoated gratings. *J. Opt. Soc. Am. A*, 8(5):763–777, May 1991.
- [18] Kofi Edee, Gérard Granet, and Jean-Pierre Plumey. Complex coordinate implementation in the curvilinear coordinate method: application to plane-wave diffraction by nonperiodic rough surfaces. *J. Opt. Soc. Am. A*, 24(4):1097–1102, April 2007.
- [19] Gene H. Golub and Charles F. van Van Loan. *Matrix Computations (Johns Hopkins Studies in Mathematical Sciences)(3rd Edition)*. The Johns Hopkins University Press, 3rd edition, October 1996.
- [20] Joseph W. Goodman. *Introduction to Fourier Optics*. Roberts & Company Publishers, third edition, December 2004.
- [21] G. Granet and B. Guizal. Efficient implementation of the coupled-wave method for metallic lamellar gratings in TM polarization. *J. Opt. Soc. Am. A*, 13(5):1019–1023, May 1996.
- [22] Gérard Granet. Reformulation of the lamellar grating problem through the concept of adaptive spatial resolution. *J. Opt. Soc. Am. A*, 16(10):2510–2516, October 1999.
- [23] Davit Harutyunyan. *Adaptive vector finite element methods for the Maxwell equations*. PhD thesis, University of Twente, 2007.
- [24] John J. Hench and Zdenek Strakos. The RCWA method - A case study with open questions and perspectives of algebraic computations. *Electronic Transactions on Numerical Analysis*, 31:331–357, 2008.

- 
- [25] R. Hiptmair. Multigrid Method for Maxwell's Equations. *SIAM J. Numer. Anal.*, 36(1):204–225, November 1998.
- [26] Jean P. Hugonin and Philippe Lalanne. Perfectly matched layers as nonlinear coordinate transforms: a generalized formalization. *J. Opt. Soc. Am. A*, 22(9):1844–1849, 2005.
- [27] John D. Jackson. *Classical Electrodynamics*. Wiley, third edition, August 1998.
- [28] Olaf T. A. Janssen. *Rigorous simulations of emitting and non-emitting nano-optical structures*. PhD thesis, 2010.
- [29] Jianming Jin. *The Finite Element Methods in Electromagnetics*. John Wiley & Sons, Inc., New York, 2002.
- [30] Hwi Kim, Il-Min Lee, and ByoungHo Lee. Extended scattering-matrix method for efficient full parallel implementation of rigorous coupled-wave analysis. *J. Opt. Soc. Am. A*, 24(8):2313–2327, August 2007.
- [31] K. Knop. Rigorous diffraction theory for transmission phase gratings with deep rectangular grooves. *J. Opt. Soc. Am.*, 68(9):1206–1210, September 1978.
- [32] D. Y. K. Ko and J. R. Sambles. Scattering matrix method for propagation of radiation in stratified media: attenuated total reflection studies of liquid crystals. *J. Opt. Soc. Am. A*, 5(11):1863–1866, November 1988.
- [33] K. Kumar, M. Pisarenco, M. Rudnaya, and V. Savcenco. Analysis, numerics, and optimization of algae growth. Submitted, 2011.
- [34] P. Lalanne, M. Besbes, J. P. Hugonin, S. van Haver, O. T. A. Janssen, A. M. Nugrowati, M. Xu, S. F. Pereira, H. Urbach, A. S. van de Nes, P. Bienstman, G. Granet, A. Moreau, S. Helfert, M. Sukharev, T. Seideman, F. Baida, B. Guizal, and D. van Labeke. Numerical analysis of a slit-groove diffraction problem. *Journal European Optical Society - Rapid Publications vol 2 07022*, 2, July 2007.
- [35] Philippe Lalanne and G. Michael Morris. Highly improved convergence of the coupled-wave method for TM polarization. *J. Opt. Soc. Am. A*, 13(4):779–784, April 1996.
- [36] Philippe Lalanne and Eric Silberstein. Fourier-modal methods applied to waveguide computational problems. *Opt. Lett.*, 25(15):1092–1094, August 2000.
- [37] Nicolas Lantos and Frédéric Nataf. Perfectly matched layers for the heat and advection-diffusion equations. *Journal of Computational Physics*, 229(24):9042–9052, December 2010.
- [38] H. J. Levinson. Is there light at the end of the road for optical lithography? *IEEE Circuits and Devices Magazine*, 18(4):50–58, July 2002.



- 
- [39] Lifeng Li. Multilayer modal method for diffraction gratings of arbitrary profile, depth, and permittivity. *J. Opt. Soc. Am. A*, 10(12):2581–2591, December 1993.
- [40] Lifeng Li. Formulation and comparison of two recursive matrix algorithms for modeling layered diffraction gratings. *J. Opt. Soc. Am. A*, 13(5):1024–1035, May 1996.
- [41] Lifeng Li. Use of Fourier series in the analysis of discontinuous periodic structures. *J. Opt. Soc. Am. A*, 13(9):1870–1876, 1996.
- [42] Peijun Li. Coupling of finite element and boundary integral methods for electromagnetic scattering in a two-layered medium. *Journal of Computational Physics*, 229(2):481–497, January 2010.
- [43] Joseph Mautz. *Integral Equation Methods for Electromagnetics (Artech House Antenna Library)*. Artech House Publishers, October 1991.
- [44] K. A. Michalski and D. Zheng. Electromagnetic scattering and radiation by surfaces of arbitrary shape in layered media. I - Theory. II - Implementation and results for contiguous half-spaces. *IEEE Transactions on Antennas and Propagation*, 38:335–352, March 1990.
- [45] M. G. Moharam and T. K. Gaylord. Rigorous coupled-wave analysis of planar-grating diffraction. *J. Opt. Soc. Am.*, 71(7):811–818, July 1981.
- [46] M. G. Moharam and T. K. Gaylord. Diffraction analysis of dielectric surface-relief gratings. *J. Opt. Soc. Am.*, 72(10):1385–1392, October 1982.
- [47] M. G. Moharam, Eric B. Grann, Drew A. Pommet, and T. K. Gaylord. Formulation for stable and efficient implementation of the rigorous coupled-wave analysis of binary gratings. *J. Opt. Soc. Am. A*, 12(5):1068–1076, May 1995.
- [48] M. G. Moharam, Drew A. Pommet, Eric 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(5):1077–1086, May 1995.
- [49] Cleve Moler and Charles Van Loan. Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later. *SIAM Review*, 45(1):3–49, 2003.
- [50] Peter Monk. A finite element method for approximating the time-harmonic Maxwell equations. *Numerische Mathematik*, 63(1):243–261, December 1992.
- [51] Peter Monk. *Finite Element Methods for Maxwell's Equations*. Oxford University Press, Oxford, 2003.
- [52] R. H. Morf. Exponentially convergent and numerically efficient solution of Maxwell's equations for lamellar gratings. *J. Opt. Soc. Am. A*, 12(5):1043–1056, May 1995.

- 
- [53] J. C. Nedelec. Mixed finite elements in R3. *Numerische Mathematik*, 35(3):315–341, September 1980.
- [54] Michael L. Overton. *Numerical Computing with IEEE Floating Point Arithmetic*. Soc for Industrial & Applied Math, 1st edition, 2001.
- [55] D. M. Pai and K. A. Awada. Analysis of dielectric gratings of arbitrary profiles and thicknesses. *J. Opt. Soc. Am. A*, 8(5):755–762, May 1991.
- [56] R. Petit. *Electromagnetic Theory of Gratings (Topics in Applied Physics)*. Springer, December 1980.
- [57] P. G. Petropoulos. An analytical study of the discrete perfectly matched layer for the time-domain maxwell equations in cylindrical coordinates. *IEEE Transactions on Antennas and Propagation*, 51(7):1671–1675, July 2003.
- [58] Maxim Pisarenco, Joseph Maubach, Irwan Setija, and Robert Mattheij. A Numerical Method for the Solution of Time-Harmonic Maxwell Equations for Two-Dimensional Scatterers. *AIP Conference Proceedings*, 1281(1):2049–2052, 2010.
- [59] Maxim Pisarenco, Joseph Maubach, Irwan Setija, and Robert Mattheij. Aperiodic Fourier modal method in contrast-field formulation for simulation of scattering from finite structures. *J. Opt. Soc. Am. A*, 27(11):2423–2431, November 2010.
- [60] Maxim Pisarenco, Joseph Maubach, Irwan Setija, and Robert Mattheij. An extended Fourier modal method for plane-wave scattering from finite structures. volume 7717, pages 77171H+. SPIE, 2010.
- [61] Maxim Pisarenco, Joseph Maubach, Irwan Setija, and Robert Mattheij. The Fourier modal method for aperiodic structures. CASA Report 10-21, February 2010.
- [62] Maxim Pisarenco, Joseph Maubach, Irwan Setija, and Robert Mattheij. Efficient solution of Maxwell’s equations for geometries with repeating patterns by an exchange of discretization directions in the aperiodic Fourier modal method. To be submitted, 2011.
- [63] Maxim Pisarenco, Joseph Maubach, Irwan Setija, and Robert Mattheij. Modified S-matrix algorithm for the aperiodic Fourier modal method in contrast-field formulation. *Journal of the Optical Society of America*, 28(7), July 2011.
- [64] Maxim Pisarenco, Bas van der Linden, Arris Tijsseling, Emmanuel Ory, and Jacques Dam. Friction Factor Estimation for Turbulent Flows in Corrugated Pipes with Rough Walls. *Journal of Offshore Mechanics and Arctic Engineering*, 133(1):011101+, 2011.
- [65] Evgeni Popov and Michel Nevière. Grating theory: new equations in Fourier space leading to fast converging results for TM polarization. *J. Opt. Soc. Am. A*, 17(10):1773–1784, October 2000.

- [66] Evgeny Popov and Michel Nevière. Maxwell equations in Fourier space: fast-converging formulation for diffraction by arbitrary shaped, periodic, anisotropic media. *J. Opt. Soc. Am. A*, 18(11):2886–2894, November 2001.
- [67] Evgeny Popov, Michel Nevière, Boris Gralak, and Gérard Tayeb. Staircase approximation validity for arbitrary-shaped gratings. *J. Opt. Soc. Am. A*, 19(1):33–42, January 2002.
- [68] Dennis W. Prather, Mark S. Mirotznik, and Joseph N. Mait. Boundary integral methods applied to the analysis of diffractive optical elements. *J. Opt. Soc. Am. A*, 14(1):34–43, January 1997.
- [69] Dennis W. Prather and Shouyuan Shi. Formulation and application of the finite-difference time-domain method for the analysis of axially symmetric diffractive optical elements. *J. Opt. Soc. Am. A*, 16(5):1131–1142, May 1999.
- [70] S. Rafler, P. Götz, M. Petschow, T. Schuster, K. Frenner, and W. Osten. Investigation of methods to set up the normal vector field for the differential method. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 6995 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, May 2008.
- [71] R. Redheffer. *Difference equations and functional equations in transmission-line theory*, chapter 12, pages 282–337. McGraw-Hill, New York, 1961.
- [72] Yousef Saad. *Iterative Methods for Sparse Linear Systems, Second Edition*. Society for Industrial and Applied Mathematics, 2 edition, April 2003.
- [73] Jyrki Saarinen, Eero Noponen, and Jari P. Turunen. Guided-mode resonance filters of finite aperture. *Optical Engineering*, 34(9):2560–2566, 1995.
- [74] Sheppard Salon and Chari. *Numerical Methods in Electromagnetism*. Academic Press, 1 edition, November 1999.
- [75] Valeriu Savcenko, Kundan Kumar, Maxim Pisarenko, Maria Rudnaya, and Sudhir Srivastava. Shape reconstruction techniques for optical sectioning of arbitrary objects. *Mathematics-in-Industry Case Studies*, 3:19–36, 2011.
- [76] A. Schadle, L. Zschiedrich, S. Burger, R. Klose, and F. Schmidt. Domain decomposition method for Maxwell’s equations: Scattering off periodic structures. *Journal of Computational Physics*, 226(1):477–493, September 2007.
- [77] S. Schot. Eighty years of Sommerfeld’s radiation condition. *Historia Mathematica*, 19(4):385–401, November 1992.
- [78] Thomas Schuster, Johannes Ruoff, Norbert Kerwien, Stephan Rafler, and Wolfgang Osten. Normal vector method for convergence improvement using the RCWA for crossed gratings. *J. Opt. Soc. Am. A*, 24(9):2880–2890, September 2007.

- [79] Eric Silberstein, Philippe Lalanne, Jean-Paul Hugonin, and Qing Cao. Use of grating theories in integrated optics. *J. Opt. Soc. Am. A*, 18(11):2865–2875, November 2001.
- [80] Pavel Solin. *Partial Differential Equations and the Finite Element Method*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2006.
- [81] Arnold Sommerfeld. *Partial Differential Equations in Physics (Pure and Applied Mathematics: A Series of Monographs and Textbooks, Vol. 1)*. Academic Press, 1949.
- [82] Randall J. Swift and Stephen A. Wirkus. *A Course in Ordinary Differential Equations*. CRC Press, 2006.
- [83] Allen Taflove and Susan C. Hagness. *Computational Electrodynamics: The Finite-Difference Time-Domain Method, Third Edition*. Artech House Publishers, 3 edition, June 2005.
- [84] Eng L. Tan. Note on formulation of the enhanced scattering- (transmittance-) matrix approach. *J. Opt. Soc. Am. A*, 19(6):1157–1161, June 2002.
- [85] A. Thornton, T. Weinhart, O. Bokhove, B. Zhang, Sar, K. Kumar, M. Pisarenco, M. Rudnaya, V. Savcenko, J. Rademacher, J. Zijlstra, A. Szabelska, J. Zyprych, Schans, V. Timperio, and F. Veerman. Modeling and optimization of algae growth. In *Proceedings of the 72nd European Study Group Mathematics with Industry (SWI 2010, Amsterdam, The Netherlands, January 25-29, 2010)*, pages 54–85. CWI, 2010.
- [86] Nico van der Aa. *Sensitivity analysis for grating reconstruction*. PhD thesis, Eindhoven University of Technology, November 2007.
- [87] B. J. van der Linden, M. Pisarenco, E. Ory, J. A. M. Dam, and A. S. Tijsseling. Efficient computation of three-dimensional flow in helically corrugated hoses including swirl. In *Proceedings of ASME Pressure Vessels and Piping Division Conference (Prague, Czech Republic, July 26-30, 2009)*, pages 655–664. ASME, July 2009.
- [88] Mark van Kraaij. *Forward diffraction modelling: analysis and application to grating reconstruction*. PhD thesis, Eindhoven University of Technology, March 2011.
- [89] J. R. Wait. Scattering of a Plane Wave from a Circular Dielectric Cylinder at Oblique Incidence. *Canadian Journal of Physics*, 33:189–195, 1955.
- [90] Zhonghua Wu and Jiayuan Fang. High-performance PML algorithms. *IEEE Microwave and Guided Wave Letters*, 6(9):335–337, September 1996.
- [91] Kane Yee. Numerical solution of initial boundary value problems involving maxwell's equations in isotropic media. *IEEE Transactions on Antennas and Propagation*, 14(3):302–307, May 1966.
- [92] Pochi Yeh. *Optical Waves in Layered Media (Wiley Series in Pure and Applied Optics)*. Wiley-Interscience, 2nd edition, March 2005.

- [93] L. Zschiedrich, S. Burger, B. Kettner, and F. Schmidt. Advanced Finite Element Method for Nano-Resonators, January 2006.
- [94] L. Zschiedrich, S. Burger, A. Schaedle, and F. Schmidt. A Rigorous Finite-Element Domain Decomposition Method for Electromagnetic Near Field Simulations, March 2008.

# Index

- adaptive meshing, 22
- adaptive spatial resolution, 23, 28
- alternative discretization, 80
- Ampère's law, 10
- angular temporal frequency, 13
- aperiodic model problem, 29
- azimuthal angle, 15
  
- background field, 28, 36
- background multilayer, 15
- background permittivity, 15
- background problem, 28
- Bloch condition, 19
- boundary element methods, 25
  
- CFL condition, 22
- complex-valued relative permittivity, 14
- conductivity, 10
- conical incidence, 18
- constitutive relations
  - time-dependent, 10
  - time-harmonic, 14
- contrast field, 28, 36
- contrast permittivity, 15
- contrast-field formulation, 54
- Courant-Friedrichs-Lewy condition, 22
- cumulative scattering matrix, 70, 72
- cumulative source vector, 72
  
- die, 2
- diffraction limit, 4
- domain decomposition, 22
  
- edge elements, 22
- electric
  - charge density, 9
  - current density, 9
  - displacement, 9
  - field, 9
- electric permittivity, 10
- electronic chip, 1
- enhanced transmittance matrix approach, 27, 68
- external current, 10
  
- Faraday's law, 9
- finite element method, 22
- finite-difference time-domain, 21
- Floquet condition, 19
- Fourier modal method, 23
- Fredholm equation of the second kind, 25
- full-matrix approach, 42
  
- Gauss's law for electric fields, 10
- Gauss's law for magnetic fields, 10
- gratings, 3
- Green's function, 24
  
- horizontal problem, 82
  
- in-die metrology, 5
- integral equation method, 24
- integrated circuit, 1
- inverse decay length, 55
  
- kernel, 25
  
- Li rules, 28
- linear time-invariant media, 10
- lithography system, 2
- local periodicity, 6

- local scattering matrix, 72
- local source vector, 72
- magnetic
  - field, 9
  - induction, 9
  - permeability, 10
- Maxwell equations
  - time-dependent, in differential form, 9
  - time-dependent, in integral form, 11
  - time-harmonic, in differential form, 13
- memory use factor, 99
- multigrid solution, 22
- non-homogeneous Redheffer star product, 90
- non-homogeneous S-matrix algorithm, 73
- non-magnetic, 14
- normal vector fields, 23
- perfectly matched layers, 28
- periodic model problem, 29
- planar incidence, 17
- plane of incidence, 17
- polar angle, 15
- polarization
  - TE, transverse electric, 17
  - TM, transverse magnetic, 17
- polarization angle, 15
- pseudo-periodic boundary condition, 18
- R-matrix algorithms, 68
- Redheffer notation, 88
- Redheffer star product, 90
- refractive index, 14
- S-matrix algorithm, 68, 69
- scribe lane, 3
- semiconducting material, 1
- Silver-Müller radiation condition, 19
- Sommerfeld radiation condition, 19
- source-free, 14
- speed-up factor, 98
- stratified medium, 15
- subnormal numbers, 98
- supercell approach, 28
- surface charge density, 12
- surface integral methods, 25
- T-matrix algorithm, 69
- the double-curl equation, 22
- time-harmonic, 13
- transistors, 1
- underflow, 98
- vacuum wavenumber, 14
- vertical problem, 82
- volume integral methods, 25
- Yee scheme, 21

# Summary

The Fourier modal method (FMM) is widely used in the diffractive optics community as an efficient tool for simulating scattering from infinitely periodic gratings. In reality the gratings are finite in size, and in applications such as lithography, it is desirable to make them as small as possible. At a certain point the assumption on infinite periodicity loses its validity. In this thesis we extend the application area of the FMM to finite structures and address the issues of stability and efficiency of the newly developed method.

The aperiodic Fourier modal method in contrast-field formulation (AFMM-CFF) is developed by placing perfectly matched layers at the lateral sides of the computational domain and reformulating the governing equations in terms of a contrast field which does not contain the incoming field. Due to the reformulation, the homogeneous system of second-order ordinary differential equations from the original FMM becomes non-homogeneous. Its solution is derived analytically and used in the established FMM framework. The technique is first demonstrated on a model problem (planar scattering of TE-polarized light by a single rectangular line). Subsequently the method is generalized to arbitrary shapes of scatterers and conical incidence.

The contrast-field formulation of the equations modifies the structure of the resulting linear systems and makes the direct application of available stable recursion algorithms impossible. We adapt the well-known S-matrix algorithm for use with the AFMM-CFF. To this end stable recursive relations are derived for the new type of linear systems. The stability of the algorithm is confirmed by numerical results.

The efficiency of the AFMM-CFF is improved by exchanging the discretization directions. Classically, spectral discretization is used in the finite periodic direction and spatial discretization in the normal direction. In the light of the fact that the structures of interest have a large width-to-height ratio and that the two discretization techniques have different computational complexities, we propose exchanging the discretization directions. This step requires a projection of the background field on the new basis introduced by the alternative discretization. For scatterers with locally repeating patterns, such as finite gratings, exchanging the discretization directions facilitates the reuse of results of previ-



ous computations, thus making the method even more efficient. As shown by numerical experiments a considerable reduction of the computational costs can be achieved.

# Samenvatting

De Fourier mode methode (FMM) wordt veel gebruikt in diffractieve optica om verstrooiing van oneindig periodieke rasters te simuleren. In werkelijkheid hebben rasters een eindige afmeting en in toepassingen zoals lithografie is het zelfs wenselijk om deze rasters zo klein mogelijk te maken. Bij een eindig raster kan daarom geen periodiciteit meer aangenomen worden. In dit proefschrift breiden we de toepassing van de FMM uit naar oneindige structuren en leiden we een stabiele en efficiënte oplosmethode af.

Een aperiodieke Fourier mode methode in contrastveld formulering (AFMM-CFF) kan worden gevonden door het plaatsen van een zogenaamde perfect aangepaste laag (PML) aan weerszijden van het rekendomein en de vergelijkingen te herformuleren in termen van contrastveld die geen componenten van het invallende veld bevat. Door deze herformulering verandert het homogene stelsel van tweede orde (gewone) differentiaalvergelijkingen dat optreedt bij FMM in een niet-homogeen stelsel. De oplossing hiervan wordt analytisch afgeleid en gebruikt in het bestaande FMM kader. Deze techniek wordt eerst toegepast op een model probleem (planaire verstrooiing van TE-gepolariseerd licht aan een enkele rechthoekige lijn). Vervolgens wordt de methode gegeneraliseerd voor willekeurige verstrooiende vormen en conische hoeken van inval.

De contrastveld formulering van de vergelijkingen verandert de structuur van het uiteindelijke lineaire stelsel en maakt de directe toepassing van bestaande stabiele recursieve algoritmen onmogelijk. In dit proefschrift wordt daarom de S-matrix algoritme aangepast voor gebruik binnen de AFMM-CFF. Voor de hierbij optredende nieuwe type lineaire stelsels worden stabiele recursieve relaties afgeleid. De stabiliteit van de algoritme wordt aan de hand van numerieke resultaten gedemonstreerd.

De efficiëntie van de AFMM-CFF wordt verbeterd door de discretisatierichtingen te wisselen. Gewoonlijk wordt een spectrale ontbinding gebruikt in de (eindige) periodieke richting en een ruimtediscretisatie in de richting loodrecht daarop. Gezien het feit dat de te onderzoeken structuren een grote breedte-hoogte verhouding hebben en dat de beide discretisatietechnieken verschillende rekenkosten hebben stelen we voor de discretisatierichtingen te wisselen. Deze stap vereist een projectie van het achtergrondveld op de

nieuwe basis die hoort bij de alternatieve discretisatie. Voor verstrooiers met zich herhalende patronen zoals eindige rasters, blijkt het verwisselen van discretisatierichtingen het hergebruik van resultaten van eerdere berekeningen op een eenvoudige manier mogelijk te maken waardoor de methode erg efficiënt wordt. Hiermee blijkt een aanzienlijke reductie van rekenkosten te worden bereikt. Dit wordt gedemonstreed aan een aantal voorbeelden.

# Curriculum vitae

Maxim Pisarenco was born in Chişinău, Moldova, on November 23rd 1983. After finishing his pre-university education at the Gaudeamus high-school (Chişinău) in 2001, he started his bachelor studies in Control Theory and Informatics at the Technical University of Moldova. In 2005 Maxim won the scholarship of the European Commission for master studies which has allowed him to obtain a double degree in Industrial Mathematics from the Eindhoven University of Technology in the Netherlands and the University of Kaiserslautern in Germany. He received his Master's degree *cum laude* in 2007. The thesis was written under the supervision of prof. dr. R.M.M. Mattheij, prof. dr. A. Klar and dr. A.S. Tijsseling, and was entitled "Friction factor estimation for flow in corrugated pipes with rough walls".

From October 2007 till September 2011 he has been working as a PhD researcher at the Eindhoven University of Technology within the Centre for Analysis, Scientific computing and Applications (CASA) under the supervision of prof. dr. R.M.M. Mattheij, dr. J.M.L. Maubach and dr. I.D. Setija. The PhD project was sponsored and in close cooperation with ASML, world's largest supplier of photolithography systems for the semiconductor industry.



