A domain integral equation approach for simulating two dimensional transverse electric scattering in a layered medium with a Gabor frame discretization.

Citation for published version (APA):

Document license:
CC BY-NC-ND

DOI:
10.1016/j.jcp.2017.05.034

Document status and date:
Published: 15/09/2017

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

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.
A domain integral equation approach for simulating two dimensional transverse electric scattering in a layered medium with a Gabor frame discretization

R.J. Dilz *, M.C. van Beurden

Gebouw Flux, Postbus 513, 5600 MB, Eindhoven, The Netherlands

**ARTICLE INFO**

Article history:
Received 17 October 2016
Received in revised form 16 May 2017
Accepted 17 May 2017
Available online 22 May 2017

**Keywords:**
Gabor frame
Integral equation
Electromagnetic scattering
Green function

**ABSTRACT**

We solve the 2D transverse-electrically polarized domain-integral equation in a layered background medium by applying a Gabor frame as a projection method. This algorithm employs both a spatial and a spectral discretization of the electric field and the contrast current in the direction of the layer extent. In the spectral domain we use a representation on the complex plane that avoids the poles and branchcuts found in the Green function. Because of the special choice of the complex-plane path in the spectral domain and because of the choice to use a Gabor frame to represent functions on this path, fast algorithms based on FFTs are available to transform to and from the spectral domain, yielding an $O(N \log N)$ scaling in computation time.

© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

For several applications in electrical engineering it is vital to have fast and accurate models to calculate the scattering of electromagnetic waves from dielectric structures of finite size. Among these are metrology for integrated circuit production [1,2], various elements on nanophotonic chips [3,4] and metamaterials [5]. For these applications, structures are often embedded in a host medium with multiple layers of different materials.

Many numerical methods have already been developed for the characterization of electromagnetic scattering in a multilayered medium, e.g. local formulations such as finite difference time domain (FDTD) [6] and the Finite Element Method (FEM) [7,8]. Global formulations that employ a Green function exist in both a time-domain formulation and a time-harmonic formulation. In the time domain the Green function can be generalized to multilayered media as well [9–11]. Since we are interested in scattering from monochromatic light-sources, we are interested in a time-harmonic formulation. Such an integral formulation requires solving a nonhomogeneous matrix equation, which can be solved efficiently with an iterative solver, especially when the matrix vector product can be computed rapidly. One popular approach is to speed up this matrix vector product by decomposing the Green function into long-range and short-range interactions, combined with a hierarchical division of the simulation domain. Examples of such methods for a homogeneous medium are the Fast Multipole Method (FMM) [12] and the Fast Inhomogeneous Plane Wave Algorithm (FIPWA) [13,14]. Extensions to multilayer media exists both for FMM [15] and for FIPWA [16]. Another popular approach for fast matrix vector products exploits the
observation that the Green function in a layered medium exhibits a translation symmetry in the direction parallel to the layer interfaces. We will focus on a method that exploits this translation symmetry.

A domain integral formulation consists of two parts. The first part is the calculation of the contrast current density from the electric field and the contrast function. The second part is an integral over the product of the Green function and the contrast current density that yields the scattered electric field. For homogeneous media, the spatial Green function is readily obtained. Among the free-space methods that employ the Green function in the spatial domain and exploit the translation symmetry are the Conjugate Gradient Fast Fourier Transform CGFFT [17] and its enhancements, such as the Adaptive Integral Method (AIM) [18] and the pre-corrected FFT [19].

To use similar methods in stratified media, the multi-layered Green function is required. The main differences between the two-dimensional free-space Green function and the two-dimensional stratified-medium Green function are the reflections and transmissions at the layer interfaces in layered media. For stratified media, an exact expression for the Green function in the spectral domain can be derived, so in principle it is possible to calculate the Green function completely in the spatial domain via a Fourier transform. However, calculating the pertaining Fourier integral is far from trivial, since there are branchcuts and poles present in the Green function. Several methods exist to calculate these so-called Sommerfeld integrals e.g. the Discrete Complex Image Method (DCIM) [20], the steepest descent path (SDP) [21], Sommerfeld tail extrapolation [22], and a method based on a perfectly matched layer [23,24].

An alternative approach is to consider the so-called spectral methods, in which the exact spectral-domain Green function is employed directly. For periodically repeating scattering structures, such as optical gratings, several spectral methods have already been developed. Important examples are the Rigorous Coupled Wave Analysis (RCWA), also called the Fourier Modal Method [25,26], and some periodic-volume-integral-equation-based methods (PVIE) [27]. Then, the periodicity can be exploited because the periodicity implies a discrete spectral domain and therefore an obvious and well-performing discretization of the spectral domain exists. These methods can be adapted to solve a-periodic structures as well, for example via perfectly matched layers (PML) [28] or supercell techniques, but even then these solvers are in essence periodic.

Here we present a mixed spatial–spectral method that is completely a-periodic in nature. This also implies that the spectral domain is now continuous instead of discrete. Consequently, branchcuts and poles in the spectral Green function need to be treated carefully. We demonstrate an approach that employs a representation of the fields in the spectral-domain complex plane, that avoids the poles found in the Green function. This representation has been specifically chosen such that the Green functions consist of smooth functions with an effectively limited support, while still allowing for efficient transformations to and from the spectral domain with $O(N \log N)$ computational complexity, for $N$ degrees of freedom. The finite support in both the spectral and the spatial domain allows for a convenient discretization in terms of Gabor frames. Owing to the Gabor frames, all operations are of $O(N \log N)$ complexity or less, thus yielding an $O(N \log N)$ scaling with the number of unknowns.

We start this paper with details about the formulation, after which we present the discretization scheme. Subsequently, the spectral complex-plane path and the representation of the Green functions on this path are illustrated. We conclude with three numerical examples to demonstrate the proposed scheme.

2. Formulation

2.1. Problem definition

Consider a layered medium, i.e. a structure of $N - 1$ horizontal layers stacked in the $z$-direction, with relative permittivities $\varepsilon_{r,n}$ and thicknesses $d_n$. The space below the stack has permittivity $\varepsilon_{r,N}$ and above the stack has permittivity $\varepsilon_{r,0}$. In layer $i$ a two dimensional dielectric object in the $x-z$ plane is described by the permittivity function $\varepsilon_r(x,z)$. The object is contained within the rectangle $x \in [-W, W]$, $z \in [z_{\text{min}}, z_{\text{max}}]$, which we call the simulation domain. Fig. 1(a) shows our scattering setup for $N = 3$ and $i = 1$. 

![Fig. 1. (a) The scattering setup. (b) The source of reflections including the definition of the different K waves.](image)
Due to the two-dimensional configuration and the polarization of the incoming field, only the \( y \) component of the electric field is nonzero, which turns our problem into a scalar problem \( \mathbf{E}(x, z) = \mathbf{y}E(x, z) \). Putting this into the Maxwell equations [29] with time convention \( \exp(j\omega t) \) yields a second order differential equation.

We make a distinction between the incident electric field \( E^i \), which solves the problem in absence of the scattering object, and the field scattered by the object \( E^s \). The combination of these fields yields the total electric field \( E = E^i + E^s \) in the simulation domain.

The contrast function in layer \( i \) is defined by

\[
\chi(x, z) = \frac{\epsilon_r(x, z)}{\epsilon_r,i} - 1.
\]

The contrast current density can be obtained through \( J(x, z) = j\omega\epsilon_0\sigma_i \chi(x, z)E(x, z) \). Similarly \( J^i(x, z) = j\omega\epsilon_0\sigma_i \chi(x, z)E^i(x, z) \) and \( J^s(x, z) = j\omega\epsilon_0\sigma_i \chi(x, z)E^s(x, z) \) are obtained from the incident and scattered part of the electric field.

2.2. The homogeneous medium Green function

With the Green function the field radiated by a contrast current can be calculated. The Green function will be represented in the spectral domain in the \( x \)-direction. The spectral domain is defined through the Fourier transformation

\[
\varphi(k_x) = \mathcal{F}_x[\varphi(x)](k_x) = \int_{-\infty}^{\infty} dx \varphi(x)e^{-jk_xx} \tag{2}
\]

and its inverse \( \mathcal{F}^{-1} \). We use the variable \( k_x \) for the spectral domain. Whenever a function has \( x \) as its argument the spatial version of the function is meant, when \( k_x \) is used as an argument its Fourier transform is meant.

We will start working on the problem within one layer, i.e. we assume a homogeneous dielectric constant in the background layer. By adding reflection and transmission coefficients we will turn this into the solution for a multi-layered problem in Section 2.3. In a homogeneous medium the Green function, \( G^h \), for a contrast current density \( J \) can be written as

\[
G^h(k_x, z|z') = \frac{e^{-\gamma|z'-z|}}{2\gamma} \tag{3}
\]

where \( \gamma = \sqrt{k_x^2 - k_y^2} \) and \( k_y = \omega\sqrt{\epsilon_{0}\sigma_{i}} \) with standard branchcut in the square root. Again \( i \) indicates the layer in which the simulation domain is located, and the \( h \) subscript will indicate the homogeneous part in this context.

We can calculate the scattered electric field \( E^s \) from current source \( J \) (not yet taking reflections into account) by

\[
E^s(k_x, z) = \int_{z_{\text{min}}}^{z} dz' \frac{e^{-\gamma(z'-z)}}{2j\omega\epsilon_{0}\sigma_{i}}J(k_x, z') + \int_{z}^{z_{\text{max}}} dz' \frac{e^{-\gamma(z'-z)}}{2j\omega\epsilon_{0}\sigma_{i}}J(k_x, z') = K^{h,d}(k_x, z) + K^{h,u}(k_x, z). \tag{4}
\]

Here we recognize that the factor \( 1/2j\omega\epsilon_{0}\sigma_{i}\gamma \) factor calculates the spectral electric field from the spectral current density at the same altitude and that the factor \( e^{-\gamma z} \) propagates this over an altitude displacement \( z \) in the \( z \)-direction. We will call \( e^{-\gamma z} \) the propagation function. The first term in Eq. (4) represents a wave moving down from the source, which we will denote by \( K^{h,d} \), and the second term a wave moving up denoted by \( K^{h,u} \).

2.3. Reflections from layer interfaces

In a multi-layered medium we have to include reflections from the layer interfaces to arrive at the complete scattered field \( E^s \) from a current source. With the use of [30, Chapter 2] and [31, Chapter 5], we can calculate the reflection coefficient from the stack of layers above layer \( i \), \( R^d(k_x) \) and from the stack below layer \( i \), \( R^u(k_x) \).

From Fig. 1(b), we deduce how the scattered field \( E^s \) can be constructed from the homogeneous field \( E^h \) together with a sum of all reflections. For the upward directed part of \( E^s \), i.e. \( K^{s,u} \), two contributions can be identified. First, there is the homogeneous part, indicated by \( \bigcirc \) in Fig. 1(b) and denoted by \( K^{h,u}(k_x, z) \). Second, there is a set of reflections that propagates upward from the layer interface below at \( z_{i+1} \) \( \bigcirc \) to \( \bigcirc \). Now we write the sum of all upward-directed reflections \( \bigcirc \) as a single effective reflection coefficient \( R^{\text{eff},u} \), multiplied by a single effective downward directed wave \( K^{\text{eff},d} \) \( \bigcirc \) and \( \bigcirc \). The sum of all reflections can then be propagated through the medium using the propagation function \( e^{-\gamma(z_{i+1}-z)} \), yielding

\[
K^{s,u}(k_x, z) = K^{h,u}(k_x, z) + e^{-\gamma(z_{i+1}-z)}R^{\text{eff},u}(k_x)K^{\text{eff},d}(k_x, z_{i+1}). \tag{5}
\]

Now the effective downward directed wave, \( K^{\text{eff},d} \) is defined as the sum of the homogeneous downward directed wave \( \bigcirc \), and the reflection of the upward-directed wave \( \bigcirc \), i.e.
\[ K^{\text{eff},d}(k_x, z_i) = K^{h,d}(k_x, z_{i+1}) + R^d(k_x)e^{-\gamma d_i} K^{h,u}(k_x, z_i). \]  

The reflection generated by the effective downward-directed \( K^{\text{eff},d} \) reflects upward with reflection coefficient \( R^u \). Thereafter it can bounce back and forth several times between the layer interfaces at \( z_i \) and \( z_{i+1} \), where it propagates a distance \( 2d_i \) between each bounce and is reflected with both \( R^u \) and \( R^d \). The first bounce is indicated by (7) and (8). This behavior is summarized in a single effective reflection coefficient \( R^{\text{eff},u} \) as

\[ R^{\text{eff},u} = R^u(k_x) \sum_{n=0}^{\infty} \left( R^u(k_x)R^d(k_x)e^{-2\gamma d_i} \right)^n. \]

This sum can be calculated using the geometric series [32, Section 2.1]

\[ R^{\text{eff},u}(k_x) = \frac{R^u(k_x)}{1 - R^u(k_x)R^d(k_x)e^{-2\gamma d_i}}. \]

Similarly, the down-directed wave is defined through

\[ K^{\text{eff},d}(k_x, z) = K^{h,d}(k_x, z) + e^{-\gamma(z-z_0)} R^{\text{eff},d}(k_x)K^{\text{eff},u}(k_x, z_i) \]

\[ R^{\text{eff},d}(k_x) = \frac{R^d(k_x)}{1 - R^d(k_x)R^u(k_x)e^{-2\gamma d_i}} \]

\[ R^{\text{eff},u}(k_x, z_i) = K^{h,u}(k_x, z_i) + R^u(k_x)e^{-\gamma d_i} K^{h,d}(k_x, z_{i+1}). \]

The sum of all upward and downward directed waves equals the scattered field \( E^s(k_x, z) = K^{\text{eff},u}(k_x, z) + K^{\text{eff},d}(k_x, z) \), which includes the homogeneous contribution of the Green function and reflections from the layer interfaces. We are now able to calculate the field including reflections as a function of a current source. When we call \( G[J](k_x, z) \) the integral operator that calculates the scattered field \( E^s \) generated by current source \( J \), we can write down the integral equation as

\[ J^s(x, z) = j\omega \varepsilon_0 \varepsilon_{r,i} \chi(x, z) E^s(x, z) = j\omega \varepsilon_0 \varepsilon_{r,i} \chi(x, z) \mathcal{F}_k^{-1} \left\{ G[J^f](k_x, z) \right\}(x, z). \]

where \( J^s \) is the unknown part of the contrast current density. Again, \( J^f \) is known and is obtained through \( J^f(x, z) = j\omega \varepsilon_0 \varepsilon_{r,i} \chi(x, z) E^f(x, z) \). The equation can be ordered with the unknown \( J^s \) on the right hand side as

\[ j\omega \varepsilon_0 \varepsilon_{r,i} \chi(x, z) \mathcal{F}_k^{-1} \left\{ G[J^f](k_x, z) \right\}(x, z) = J^s + j\omega \varepsilon_0 \varepsilon_{r,i} \chi(x, z) \mathcal{F}_k^{-1} \left\{ G[J^f](k_x, z) \right\}(x, z). \]

Note that this formulation is somewhat different from that in other papers, e.g. [33].

3. Discretization

3.1. The z direction: piecewise-linear functions

We use piecewise-linear (PWL) expansion functions as discretization in the \( z \) direction. The expansion functions are

\[ \Lambda_n(z) = \begin{cases} 1 - \frac{|z-n\Delta-z_{\text{min}}|}{\Delta} & \text{if } |z-n\Delta-z_{\text{min}}| < \Delta, \\ 0 & \text{if } |z-n\Delta-z_{\text{min}}| > \Delta. \end{cases} \]

(11)

with \( \Delta \) the stepsize in the \( z \) discretization. For testing we use Dirac-delta functions, since these lead to a well-conditioned problem [34]. We use a subscript \( n \) to indicate the basis function to which a field corresponds e.g. \( E_n(k_x) = E(k_x, z_{\text{min}} + n\Delta) \). The maximum value for \( n \) is \( N_z = (z_{\text{max}} - z_{\text{min}})/\Delta \). Now we can approximate for example the electric field using these expansion function by

\[ E(k_x, z) \approx \sum_{n=0}^{N_z} E_n(k_x) \Lambda_n(z). \]

(12)

With the use of this discretization in the \( z \) direction we can write the integral with \( G^h \) in Eq. (4) in more detail as

\[ K^{h,u}_n(k_x) = \int_{z_{\text{min}}}^{z_{\text{min}}+n\Delta} dz' \sum_{n'=0}^{n+1} G^h(k_x, n\Delta|z')k^2_{n'} J_{n'}(k_x) \Lambda_{n'}(z'). \]

We use a recursive algorithm [35] to find the result of the integral for each \( n \). We find
\[ K_{n+1}(k_x) = K_n(k_x) e^{-\gamma \Delta} - J_n(k_x) \int_0^\Delta dz' k_0^2 \Lambda_0(z') e^{-\gamma (\Delta - z')} \frac{2}{\gamma} - J_{n+1}(k_x) \int_0^\Delta dz' k_0^2 \Lambda_1(z') e^{-\gamma (\Delta - z')} \frac{2}{\gamma} \]

\[ = K_n(k_x) e^{-\gamma \Delta} + J_n(k_x) h_m^u(k_x) + J_{n+1}(k_x) h_e^u(k_x). \]

where we introduced \( h_m^u(k_x) \) and \( h_e^u(k_x) \) for the result of the integrals over \( z' \), here the \( m \) and \( e \) subscript stand for integrals to the middle or to the end of basis function \( \Lambda_n(z) \). In the other direction, \( K_{n+1}(k_x) \) is defined as

\[ K_{n-1}(k_x) = K_n(k_x) e^{-\gamma \Delta} - J_n(k_x) \int_{-\Delta}^0 dz' k_0^2 \Lambda_0(z') e^{-\gamma (z' - \Delta)} \frac{2}{\gamma} - J_{n+1}(k_x) \int_{-\Delta}^0 dz' k_0^2 \Lambda_{-1}(z') e^{-\gamma (z' - \Delta)} \frac{2}{\gamma} \]

\[ = K_n(k_x) e^{-\gamma \Delta} + J_n(k_x) h_m^d(k_x) + J_{n-1}(k_x) h_e^d(k_x). \]

The discretized scattered field \( E_n^h(k_x) \) can be calculated by adding the terms that represent the reflections in Eq. (5) to the homogeneous waves \( K_n^{h,w/d}(k_x) \).

3.2. The \( x \) direction: Gabor frames

For the discretization in the \( x \) direction we use a Gabor frame in the spatial as well as the spectral domain. The Gabor frame is defined in the exposition [36, Chapter 8]. We employ the Gaussian window function

\[ g(x) = 2^\frac{1}{2} e^{-\pi \frac{x^2}{K^2}}, \]

where \( X \) defines the width of the window function. The Gabor frame is defined as

\[ g_{mn}(x) = g(x - m\alpha X) e^{in\beta Kx}, \]

where \( K = 2\pi / X \) the spectral window distance and \( \alpha \) and \( \beta \) define the oversampling. We use rational oversampling where \( \alpha\beta = p/q \) with \( p < q \) and \( p, q \in \mathbb{N} \). In principle, the Gabor coefficients can be calculated from the dual frame \( \eta_{mn}(x) \) by

\[ f_{mn} = \int_{-\infty}^{\infty} dx f(x) \eta_{mn}(x), \]

where the dual frame is found from

\[ \eta_{mn}(x) = \eta(x - m\alpha X) e^{in\beta Kx}. \]

When there is oversampling there is a freedom of choice for the dual window \( \eta \). The dual window \( \eta \) we use to calculate Gabor coefficients is the one obtained using the Moore–Penrose inverse [36,37].

The spectral Gabor frame is simply the Fourier transform of the Gabor frame in Eq. (16). More details can be found in [38].

4. Complex-plane spectral path

4.1. Poles, branchcuts and rapid oscillations

From Eq. (9) it is clear that Fourier transformations are an essential part of the presented algorithm, since the contrast multiplication should be executed in the spatial domain and the Green function operator can be handled in the spectral domain. It is however difficult to represent the Green function \( \mathcal{G} \), for which an exact definition only exists in the spectral domain, because \( \mathcal{G} \) contains poles, branchcuts, and rapid oscillations in the spectral domain. It is therefore not trivial to represent the Green function efficiently using Gabor frames. Calculation of the Green operator in the spatial domain, through so-called Sommerfeld integrals, is possible but tedious. We would like to find an efficient representation of \( J, E \) and \( \mathcal{G} \) in the spectral domain to compute \( \mathcal{G} \) working on \( J \). We have several requirements for this representation.

1. We aim for a straightforward representation in which we can contain the complete integral operator \( \mathcal{G} \). We do not want to take the poles and branchcuts into account separately.
2. A fast transformation to the spectral-domain representation should be available for \( J \) and a fast transformation back to the spatial domain for \( E \).
3. Since a spectral multiplication corresponds to a spatial convolution, we need the representation to hold over the entire spatial domain. Otherwise, subsequent convolutions will induce errors.
4. We only need the fields in the simulation domain. A representation that holds over the entire spatial domain will be inefficient, as it carries more information than we need.

There is clearly some tension between points 3 and 4. A method to solve these issues is to represent functions not on the real spectral axis, but in the spectral complex plane.

The first challenge is that the Green operator $G$ contains branchcuts in the effective reflection coefficients starting at $k_x^2 = \omega^2 \mu_0 \varepsilon_0$ and $k_y^2 = \omega^2 \mu_0 \varepsilon_N$ toward $k_x = \pm j\infty$ as indicated in Fig. 2(a).

Additionally, the effective transmission and reflection coefficients (7) can also have poles corresponding to guided waves. In [32, Chapter 2.7] and [39] bounds are given within which the poles are located. For a lossless multi-layered medium it can be shown that the poles must have a distance of at least $k_0 = \omega / \sqrt{\mu_0 \varepsilon_0}$ from $k_x = 0$ and that they lie on the real $k_x$ axis. It was shown that for lossy media the poles also have a minimum distance away from $k_x = 0$ and can be found in solely in the northwest and southeast quadrants of the complex plane.

The last difficulty in the Green function are rapid oscillations. The propagation function $\exp(-\gamma z)$ oscillates rapidly for $-\mathrm{Re}(k_b) < k_x < \mathrm{Re}(k_b)$ at large $z$, since $\gamma = \sqrt{k_x^2 - k_0^2}$ has a dominant imaginary part on this range. It is hard to capture these oscillations with a small number of Gabor coefficients.

4.2. Functions in the complex plane

In [32, Chapter 2.7] several methods are mentioned for integration paths that avoid the poles and branchcuts to calculate Sommerfeld integrals efficiently. Since the discontinuities in the reflection coefficients are located in the northwest and southeast quadrant of the complex plane, we would like to discretize functions on a path through the southwest and northeast quadrant. We want to represent all functions in the spectral domain on such an integration path to bypass the poles and branchcuts.

The key observation is the complex shift

$$\mathcal{F}[f](k_x \pm jA) = \int_{-\infty}^{\infty} f(x) e^{-j(k_x \pm jA)x} dx = \mathcal{F}[f(x)e^{\mp Ax}](k_x).$$

We see that it is possible to generate a shift of a coordinate in the complex spectral domain of $\pm jA$, by multiplying the spatial representation by $e^{\mp Ax}$ and then carrying out a standard real-axis Fourier transform. As mentioned before, the Gabor frame allows for fast and efficient Fourier transformations and multiplications of functions.
To arrive at an integration path passing through the southeast and northwest quadrants we have to split the spectral domain integration path in pieces that we will handle separately. The path we choose can be written as the union of three line segments, parametrized by the real valued parameter \( \tau \), see Fig. 2(a), i.e.

\[
k_x(\tau) = \begin{cases} 
\tau - jA & \text{if } \tau < -A \\
(1 + j)\tau & \text{if } -A \leq \tau < A \\
\tau + jA & \text{if } \tau > A.
\end{cases}
\]  

This path is also used in [40] and the path also bears some resemblance to the steepest-decent path used in [21].

In principle, we should also integrate over the line segments \([-\infty, -\infty - jA] \) and \([\infty, \infty + jA] \) to close the contour, but asymptotically all functions of interest are zero on these intervals, so we can safely ignore them owing to Jordan’s lemma. We chose \( A \) small compared to the complete spectral range that we discretize, so there is only a little bit of information contained in the middle part of this representation. Although there is some freedom for the choice of \( A \) within which the overall algorithm performs well, we found that the choice \( AW \approx 3 \), with \( W \) the width of the simulation domain, will yield satisfactory results in general. An optimal choice for \( A \) will somewhat vary, depending on the shape of the simulation domain and the required accuracy.

We can identify three types of functions that are part of the Green function that we need to represent on this complex path. These three types are the propagation function \( \exp(-\gamma z) \), the current to field functions \( h^u_m, h^d_m, h^u_t \) and \( h^d_t \) (Eq. (13)), and the reflection coefficients \( R^{\text{eff},u} \) and \( R^{\text{eff},d} \) (Eq. (7)). As can be seen in Fig. 2(b), the propagation function is much better behaved on the complex path and the same holds for \( h^u_{m/e}(k_x) \), since these functions consists of a \( z \)-integral over the propagation function. Fig. 2(c) shows the effect of going around the poles on the complex path: the poles are smoothened as well. The obstacles mentioned in Section 4.1 have disappeared through the use of this complex spectral path.

4.3. Transforming to and from the complex spectral path

For the spectral domain representations on the outer parts in Eq. (20), i.e. \(|\tau| > A \) we use Eq. (19). We will adopt the notation \( f_1(\tau) = f(\tau - jA) \) and \( f_R(\tau) = f(\tau + jA) \) for the representation on the left and on the right part of the complex spectral path, respectively.

To calculate the spatial representation of a function from a spectral representation we employ Eq. (19) the other way around. However, here we also have to enforce the end of the domain at \( \tau = \pm A \) by means of cutoff functions \( c \) defined by

\[
c_L(\tau) = U(A - \tau) \]

\[
c_R(\tau) = U(\tau - A),
\]

with \( U \) the Heaviside step function. We can now calculate the contribution of \( f_{L/R} \) to the spatial domain by

\[
f(x) = e^{Ax} f_x^{-1} [c_L(\tau) f_L(\tau)](x) + e^{-AX} f_x^{-1} [c_R(\tau) f_R(k)](x) + \int_{-A-jA}^{A+jA} dk_x f(k_x) e^{ik_x x},
\]

where the last integral is still there since we did not yet describe the discretization of the middle part of the integral.

An important remark on the middle part of the complex integration path is that it does not contain much information. Optimization for speed is not crucial on the middle part as long as the time spent to calculate its contribution is negligible compared to the time needed for the left and right parts. The applied method only needs to be accurate enough.

When we assume that there are no poles in a range of \( \sqrt{2A} \) around 0 (otherwise we can choose a smaller value for \( A \)), we can employ a Taylor series to approximate functions in this part of the spectral domain. The Taylor series has the advantage that we can use derivatives around \( k_x = 0 \) to make a continuation into the complex plane.

To calculate the derivatives of a function \( f(k_x) \) represented by Gabor coefficients \( f_{mn} \) we define

\[
\tilde{f}_d = f^{(d)}(0) = \sum_{m,n} f_{mn} g^{(d)}_{mn}(0),
\]

where the derivatives from the spectral Gabor frame can be easily calculated from Eq. (16). To calculate the spatial contribution due to the derivatives \( \tilde{f}_d \), the Taylor series is used to approximate \( f(k_x) \) around \( k_x = 0 \). From

\[
f(k_x) = \sum_d k_x^d \frac{\tilde{f}_d}{d!}
\]

we can write the integral in Eq. (22) as
Fig. 3. Branchcuts can interfere with the attenuating parts of \( f_{L/R}(k_x) \).

\[
\int_{-A-jA}^{A+jA} dk_x f(k_x) e^{jkx} = \sum_d \frac{\hat{f}_d}{d!} i_d(x) \tag{25}
\]

where

\[
i_d(x) = \int_{-A-jA}^{A+jA} dk_x k_x^d e^{jkx}. \tag{26}
\]

The Gabor coefficients of the \( i_d(x) \) functions can be computed during the initialization phase of our algorithm. Additionally, the Taylor series of the product of two functions is needed when multiplying quantities that are represented on the spectral path, such as in Eqs. (4)–(8). This product can be obtained through the general Leibnitz rule:

\[
(\hat{f} \hat{h})_d(0) = \sum_m \frac{d!}{(d-m)!m!} \hat{f}_{d-m} \hat{h}_m. \tag{27}
\]

We typically need around 10 terms in the Taylor series for a simulation region of one wavelength in the \( z \) direction and three digits precision.

5. Approximation of functions

The functions that make up the Green operator \( G \), i.e. \( h_{m/e} \), \( R^{\text{eff}} \) and \( e^{-\gamma \Delta} \), need to be approximated accurately on the complex-plane spectral path. Additionally, in the spatial domain \( \chi \) has to be approximated using Gabor coefficients. We will now give some details on how we obtain approximations for these functions.

5.1. General remarks

We use the same methods as Bastiaans employs in [37] to calculate the Gabor coefficients for a function efficiently. To increase the accuracy we use some oversampling, i.e. we calculate Gabor coefficients for a larger range of index \( n \) in Eq. (17), and then discard the coefficients with large \( n \), which we do not need. This leads to a finer sampling of the function and therefore to a higher accuracy. Typically oversampling by a factor of four yields three digits precision.

The functions \( f_L(\tau) \) and \( f_R(\tau) \) are defined on the \( \text{Im}(k) = \pm A \) and \( \text{Re}(k_x) < -A \) and \( \text{Re}(k_x) > A \) respectively (see Eq. (20)). However, since we use limited number of Gabor coefficients, the approximation of the function can not just stop abruptly at \( \text{Re}(k_x) = \pm A \). Beyond \( \pm A \) the \( f_{L/R} \) have to be attenuated smoothly. As we illustrate in Fig. 3, the approximation on the solid line is continued for some distance along the dashed line to let \( f_{L/R} \) be smooth at \( \text{Im}(k_x) = \pm A \). Continuing the function \( f_{L/R} \) on the dashed line with an attenuating factor is not always possible, since \( \gamma \) has branchcuts located at \( k_x = \pm k \pm jA \) with \( k > 0 \), making approximations discontinuous at some point beyond \( \text{Re}(k_x) > 0, \text{Re}(k_x) < 0 \) respectively. The Gibbs ringing created by such a discontinuity significantly deteriorates the accuracy of the \( f_{L/R} \) approximation, since the Gibbs ringing carries over some distance. For this reason it is important that the functions to be approximated are made continuous along the complete attenuating region.

5.2. Propagation function

The \( \Delta \)-distance propagation function \( \exp(-\gamma \Delta) \) is multiplied by itself \( N_z \) times in during the recursion in Eq. (13). To guarantee numerical stability, this function has to be equal to or smaller than 1 in modulus everywhere. In the domains of \( f_L \) and \( f_R \) this function is well behaved, but it exhibits branchcuts at \( k_x = \pm jA \). Avoiding the branchcut by continuing on the other Riemann surface lets the function increase beyond one, as is shown in Fig. 4a.

We solve this issue by multiplying a linear continuation of \( f_{L/R} \) beyond \( k_x = \pm A/2 \pm jA \) by a Gaussian. For \( f_R \) we calculate the Gabor coefficients from the function
5.3. Reflection coefficients

For the reflection coefficients, the continuation of the function beyond \( \tau = \pm A \) is difficult, because we can expect to encounter poles in that region. However, we do not have the requirement of the function being smaller than 1, so we can use the same technique as for the propagation in Eq. (28) function but now with \( \alpha = 1 \).

Another critical point is that an analytic expression for the reflection coefficients does not exist in general. Therefore, we cannot calculate the derivatives for \( f_d \) analytically. The way we approximate the derivatives is by fitting a power series through the values of \( f \) at different \( k_m \). From this power series we can approximate the derivatives. We define \( k_m = -\sqrt{2A}\hat{m}/M, \ldots, \sqrt{2A}\hat{m}/M \) for \( m \in \{-M, -M + 1, \ldots, M - 1, M\} \), with \( 2M + 1 \) larger than the total required number of derivatives. Now we can calculate fit values \( f_m = f(k_m) \).

If the Taylor series Eq. (24) is to hold, then we enforce:

\[
  f_m = \sum_{d=0}^{2M} \frac{k_{md} \hat{f}_d}{d!} = \sum_{d=0}^{2M} K_{m,d} \hat{f}_d,
\]

which is a matrix equation on the right, with \( K_{m,d} = k_{md}/d! \). This Vandermonde system can be solved by (pseudo-)inverting the matrix, which yields the coefficients \( \hat{f}_d \). Since this Vandermonde system is small, ill conditioning is not a problem.

5.4. Cut function and contrast function

The cut functions \( c_{l,R}(k) \) in Eq. (21) and the contrast function \( \chi(x) \) Eq. (1) are the only discontinuous functions present in the numerical scheme. This means that we cannot simply use the fast Gabor transformation to calculate their coefficients, because this method requires samples of the function on an equidistant grid and can therefore not sample discontinuities accurately. In principle we would have to calculate the integrals in Eq. (17), which is challenging. The easiest way out is to use massive oversampling (by a factor of 1000 or more) with the fast algorithm of [37]. Since this function needs to be calculated only once, during initialization, its computation time is not very critical.

6. Summary of the algorithm

We will now summarize the steps that need to be carried out for the complete algorithm. We have ordered these steps in a list to emphasize the chronological order of these operations.

- **Gabor frame** Set up the Gabor frame in both the spatial and spectral domain, i.e., calculate an interpolation list of the dual window function in Eq. (18) in both the spectral and spatial domain. Initialize the multiplication operation on the spatial and spectral Gabor frame by calculating the \( A_{k,lm} \) in Eq. (26) of [38].

---

Fig. 4. (a) A plot of the propagation function \( e^{-\gamma \Delta} \) on the line \( k_\xi = k - jA \). Its absolute value is clearly larger than one for \( k < 0 \). (b) The continuation using Eq. (28) for different values of \( \alpha \). For \( \alpha = 1 \) and \( \alpha = 30 \) the absolute value of the propagation function stays below 1. However, for \( \alpha = 30 \) the function is not smooth enough for good approximation with Gabor coefficients.

\[
f_{k}^R(\tau) = \begin{cases} 
  e^{-\gamma \Delta} & \text{if } \tau > A/2 \\
  (a\tau + b)e^{-\alpha(\tau-A/2)^2} & \text{if } \tau \leq A/2, 
\end{cases}
\]

where \( a \) and \( b \) are fitted so the function has a continuous derivative at \( \tau = A/2 \) and \( \alpha \) is chosen such that this function will just be smaller than 1 for all \( k \), with \( \text{Im}(k) = \mp A \), as is shown in Fig. 4(b). When \( \alpha \) is large, the transition is very fast and the Gabor frame may need too many coefficients. When \( \alpha \) is small then \( f_{k}^R \) increases to values larger than one, potentially destabilizing our algorithm. We use a numerical optimization to find the largest \( \alpha \) with which the approximation does not increase beyond 1.
The numerical results

7.1. Accuracy

We have simulated two cases to validate our code. The first case consists of two blocks embedded in a three-layer medium as depicted on the left-hand side of Fig. 5. The second case consists of eight lines on top of a silicon substrate and is depicted on the right-hand side of Fig. 5.

The first case was validated using JCMWave software package [8], which uses a finite-element algorithm. The second case was validated using the algorithm of [28], which uses RCWA with PMLs.

For both simulations we used a Gabor frame with window width of $X = 250$ nm and $\alpha = \beta = \sqrt{2/3}$ oversampling. For the first case we used 13 spatial window functions and 25 modulation frequencies, which is a total of 325 unknowns per sample in $z$. To increase the accuracy we used a factor 1.5 oversampling in the spectral domain. In the $z$ direction we used 21 samples in $z$. This results in one unknown per 5 nm in the $z$ direction and one unknown per 8 nm in the $x$-direction. For the second testcase we used nine spatial windows, since the contrast source extends over a wider range in the $x$-direction. For case two a spectral oversampling of 1.2 was already sufficient. In Fig. 6 we show the electric field strength around the objects for both cases.

From Fig. 7, it can be clearly seen that the results from this algorithm coincide with the reference results up to a relative difference of around $10^{-3}$. Only around the edges of the blocks a somewhat larger error is observed, because the analytic Gabor frame cannot exactly represent the jump in the second derivative of the electric field.

We choose to plot the error for both cases at the edge of the blocks, since the scattered field generally behaves worse around discontinuities. However, we observed that the error remains more or less constant throughout the whole simulation domain. The difference between the reference solution and our simulation results can be tightened by increasing the sampling.
7.2. Computational efficiency

To study the performance of the proposed algorithm, we first consider the convergence of the iterative solver BiCGstab(2) [41,42], since fast convergence is critical to computational efficiency. We have used the scattering setup of the first testcase in Fig. 5, but with the dielectric constant of the objects increased to $\varepsilon_r = 30$. We show results with a higher contrast $\chi$ since a high contrast usually requires more iterations, and with a low contrast the convergence is too fast for an insightful plot. Fig. 8 shows that the residual error converges rapidly and that the convergence is not very sensitive to the discretization in the $x$ direction.

Another important aspect of this type of algorithm is the computation time. In [43], several methods have been tested for scattering problems in multilayered media. According to this article, the most competitive methods seem to be RCWA (or Fourier Modal Method) and FEM. Other methods mentioned here are FDTD implementations, which all scored poorly in accuracy, a hybrid method utilizing both FEM and RCWA, and a volume integral method (VIM) that was accurate but slower than some of the competition. Therefore, we focus on FEM and RCWA e.g. in [8,44,45].

To compare the present method to RCWA and FEM, we have first plotted timing results of the other methods in Fig. 9. On the vertical axis we put the relative computation time divided by the number of discretization points, $N_x$, so a horizontal line corresponds to an $O(N_x)$ algorithm, with $N_x$ proportional to the number of unknowns used in the $x$-direction. Some of the results are generated for periodic scatterers, which is in principle a different class of solver, however, the scaling of the computational efficiency is the same as their a-periodic counterparts that use PMLs or supercell techniques. In this graph
we clearly see RCWA is computationally more intense than \( O(N_x \log N_x) \), where \( N_x \) the number of harmonic functions in the \( x \)-direction. RCWA scales as \( O(N_x^2) \) or \( O(N_x^3) \), depending on the implementation. For FEM, there are two variables, the number of mesh elements \( N_t \) and the polynomial refinement \( N_p \). Since with FEM the discretization is in two directions we compare its performance by putting \( N_x \) equal to \( N_t = \sqrt[N_t]{N_x} \) or \( N_x = \sqrt[N_p]{N_x} \) on the horizontal axis.

We would like to emphasize that in [46] a FEM-algorithm has been published that scales linear as \( O(N_x) \) for a 3D case. Although this is somewhat superior to the \( O(N_x \log N_x) \) scaling in the present method, FEM has the downside of having to discretize the entire multilayer domain and added PMLs, whereas the present method only discretizes at positions where objects are located. There are cases where this advantage will outweigh the \( O(\log N_x) \) penalty of the present method, for example in the second testcase. For that testcase, FEM needs a region of at least 1220 nm in the \( z \)-direction to be discretized, while the present algorithm only uses 70 nm of discretized space in the \( z \)-direction.

In Fig. 10 we have plotted the computation time of the present method against the timing of an RCWA implementation [28] for the scattering case on the right-hand side of Fig. 5. We used 20 discretization steps in the \( z \) direction for the present method and 20 slices for RCWA to mimic slanted-boundaries behavior. In the case of RCWA the wavenumber of the highest harmonic function was calculated through \( 2\pi / HW \), \( H \) being the number of harmonics, from 1 to 512 and \( W = 1900 \) nm the period of the simulation domain. For the present method, we increased the spectral range by increasing the number of the \( n \)-index basis function in Eq. (16). Clearly, for RCWA the difference in computation time between TE and TM polarization is small, which justifies that we compare our method against results for TM polarization in Fig. 9. Although RCWA is noticeably faster for a small number of harmonics, this spectral range is too low to accurately capture the details of the problem. When a large spectral range is required, either for higher accuracy and/or for a larger simulation domain, the present method is clearly beneficial.

### 7.3. Application to a grating coupler

As a final example we show results for a grating coupler that was inspired by [47]. The application to this type of problem is challenging, since a grating coupler is a device that couples an incident field to a guided wave and therefore it shows the accuracy of the presented method for guided waves. The dielectric waveguide consists of a thin high-contrast layer deposited on a thick low-contrast layer. Waves are coupled into the high contrast layer through a set of grooves, as illustrated in Fig. 11. The dimensions of this grating coupler were not further optimized for this particular wavelength to the extent of what was done in [47].
Fig. 10. The computation times for the RCWA implementation in [28] compared to the present method. The cross-over point of both trend lines corresponds to a spatial detail of 7 nm, still coarser than the detail level of 5 nm in the z-direction. The present algorithm was tested on a single core of an Intel i7-4600u processor.

Fig. 11. A grating coupler for TE mode waves.

Fig. 12. The real part of the scattered electric field in the grating coupler of Fig. 11.

To characterize this setup, a Gabor frame with window width $X = 1550$ nm and $\alpha = \beta = \sqrt{2/3}$ truncated to 17 window functions with 241 modulation frequencies was used in the x-direction. In the z-direction 21 basis functions were used with a width of 3.5 nm to span the height of the grooves. The Taylor series in the spectral integration path was truncated at 33 terms. The required computation time to solve the complete problem was 560 seconds. Again, this result was validated against results generated with the FEM-algorithm JCMWave [8], which yielded a relative error of $4.6 \times 10^{-5}$ between the present algorithm and the FEM reference. Fig. 12 clearly shows that a guided wave is induced in the high-contrast layer that travels to the left.

The convergence of the accuracy against different discretization parameters is shown in Fig. 13. For each of the figures all simulated parameters are kept at the values mentioned above, except for the one in which a parameter sweep is performed. In Fig. 13(a), the convergence is shown against the number of modulation frequencies in the Gabor frame. The number of modulation frequencies ranges from 11 to 241, which corresponds to a sampling frequency of 120 nm down to 5.2 nm. In Fig. 13(b), the convergence is shown against the number of PWL functions that discretize the z-direction in the range from 2 to 21. This corresponds to a width $\Delta$ of the PWL function, Eq. (11), from 70 nm to 3.5 nm. In Fig. 13(c), the convergence is shown against the number of terms in the Taylor series Eq. (25). Since the calculation time is not significantly affected by the number of terms, the safest strategy is to use a rather large number of terms in the Taylor series. Clearly, a truncation
at 8 terms is already sufficient. For a large number of terms in the Taylor series, the accuracy is limited via the PWL basis and the Gabor frame.

8. Conclusion

We have presented an algorithm capable of calculating the scattering from 2D dielectric objects embedded in a layered medium, using a fully a-periodic approach under illumination of a TE-polarized wave based on a domain integral equation. The use of a Gabor frame makes it possible to efficiently approximate functions in both the spatial as well as the spectral domain simultaneously, which we exploit by carrying out the contrast multiplication in the spatial domain and the Green operator in the spectral domain.

In the spectral domain we use a representation on a path in the complex spectral plane instead of on the real axis. An advantage of this complex spectral path is that the Green operator is smoother and therefore easier to approximate. The particular choice for this path allows for a fast FFT-based transformation between the spatial representation and the complex-path spectral representation. The computation time of the resulting algorithm scales as $O(N_w \log N_w)$.

Acknowledgements

The authors would like to thank Dr. Maxim Pisarenco for providing us with reference data using the algorithm described in [28].

This work is part of the research program HTSM with project number 12786 which is (partly) financed by the Netherlands Organisation for Scientific Research (NWO).

References