Memory sparing, fast scattering formalism for rigorous diffraction modeling

W Iff\textsuperscript{1,2}, T Kämpfe\textsuperscript{1}, Y Jourlin\textsuperscript{1} and A V Tishchenko\textsuperscript{1}

\textsuperscript{1} University Jean Monnet, Laboratoire Hubert Curien UMR CNRS 5516, Bât. F, 18 Rue Benoit Lauras, F-42000, Saint-Etienne, France
\textsuperscript{2} Formerly with University of Erlangen-Nuremberg, Chair of Optics and Graduate School in Optical Technologies (SAOT), Staudtstrasse 7/B2, D-91058 Erlangen, Germany

E-mail: wolfgang.iff@univ-st-etienne.fr

Received 28 October 2016, revised 26 March 2017
Accepted for publication 28 April 2017
Published 8 June 2017

Abstract

The basics and algorithmic steps of a novel scattering formalism suited for memory sparing and fast electromagnetic calculations are presented. The formalism, called ‘S-vector algorithm’ (by analogy with the known scattering-matrix algorithm), allows the calculation of the collective scattering spectra of individual layered micro-structured scattering objects. A rigorous method of linear complexity is applied to model the scattering at individual layers; here the generalized source method (GSM) resorting to Fourier harmonics as basis functions is used as one possible method of linear complexity. The concatenation of the individual scattering events can be achieved sequentially or in parallel, both having pros and cons. The present development will largely concentrate on a consecutive approach based on the multiple reflection series. The latter will be reformulated into an implicit formalism which will be associated with an iterative solver, resulting in improved convergence. The examples will first refer to 1D grating diffraction for the sake of simplicity and intelligibility, with a final 2D application example.

Keywords: scattering theory, diffraction and scattering, S-vector algorithm, S-matrix, diffraction gratings, Fourier method

1. Introduction

There is an ongoing high demand for efficient, scattering-based rigorous diffraction simulation of planar structures (periodic \cite{1-5} as well as non-periodic \cite{3, 6, 7}), as well as non-planar structures \cite{8}. Consequently, methods efficiently calculating the collective scattering behaviour of interacting single scattering items are required \cite{9-11}. This motivates the development of efficient, scattering-based diffraction computation methods \cite{8, 10-17} and particularly of scattering formalisms \cite{18, 19}.

Scattering formalisms are needed for the rigorous diffraction modeling of complex elements, particularly for grating computation, where a widely adopted procedure is the reduction of Maxwell’s equations from a system of partial differential equations to a system of ordinary differential equations by projection onto a set of basis functions \cite{3, 4, 20}. These basis functions are of the type of forward and backward functions (up and down functions, called up and down amplitudes in this paper), decomposing the field into forward and backward propagating, as well as exponentially decreasing and increasing, fractions. Considering these forward and backward functions at certain points in the modulated region and taking into account the boundary conditions, a 1D boundary value problem is obtained. A memory sparing, fast solution of this boundary value problem for a given cluster of scattering items without sacrifice of flexibility is the subject of \cite{18, 19}, and this paper gives the rationale and the detailed description of the solution, as well as application examples. Here, ‘fast’ means a complexity linear or almost linear (\(N \ln(N)\)) with respect to the number \(N\) of considered diffraction orders.

There are various established procedures for the scattering based solution of 1D boundary value problems \cite{10, 11, 21-23}. In the case of the formulation of the 1D boundary value problem with respect to up and down
functions, recognized and popular scattering-based algorithms are the S-matrix algorithm [21, 22] and the Bremmer series method [23] or rather the ‘multiple reflection series method’ [24–26] (see below).

The S-matrix algorithm operates with reflection and transmission matrices assigned to a stack of layers and requires matrix inversions. It yields reliable results but is of \(O(N^3)\)-complexity [21, 22]. After having been used in early publications [27, 28], the S-matrix algorithm and related variants spread widely in grating computation, particularly in the nineties [21, 22, 29–31].

The Bremmer series method [23], also called ‘multiple reflection series method’ when applied to grating diffraction [24–26], uses a different approach: it converts a 1D boundary value problem into a series of initial value problems, and has an explicit, adding-up character. The value problem into a series of initial value problems, and has an explicit, adding-up character. The \(i\)th summand in the series corresponds to an \(i\)-fold reflected wave. The Bremmer series method operates with internal reflections throughout, and does not eliminate field quantities from the inner observational planes. It has first been formulated for wave propagation in a 1D inhomogeneous medium by Bremmer [23], where the author also proved that the series is a formal solution of the Helmholtz equation. Convergence criteria for the Bremmer series have been given in [32–34]. Bremmer-like series converging faster than the Bremmer series have been formulated in [35–38].

The first use of the Bremmer series in the more general Sluiter-formulation [38] applied to diffraction orders for grating diffraction calculation was published in 1991 by Pai and Awada [24] and called ‘multiple reflection series’. The propagation method used at the single scattering layers was RCWA [24]. Neviere and Montiel combined the multiple reflection series with the differential method [25, 26]. For this, they divided gratings into layers and inserted infinitesimally thin slices of a homogeneous basis medium between these layers. The diffraction computations at the single layers were conducted by the differential method. The multiple reflection series was then applied for the calculation of the scattering behaviour of the entire grating from the scattering behaviour of these single layers. Li investigated the Bremmer—or, rather, the multiple reflection series—theoretically in [39] and proved the formal equivalence between this series and the S-matrix propagation algorithm: he established the validity of the multiple reflection series as a solution of Maxwell’s equations to the grating problem.

We have realized, however, that the Bremmer series method is not an end in itself, and that the above-described development may just be the base for the origination of a highly efficient and flexible scattering formalism, implementing the following two algorithmic key processes:

- We diffract the vectors of the up and down functions at the scattering items by a fast method such as [12–16] (here: GSM [16]) instead of utilizing a slow method computing entire S-matrix blocks at \(O(N^3)\)-complexity and storing them as a preliminary measure as in [24–26]. This leads to \(O(N \ln(N))\)-complexity of the entire method.

- We derive the implicit counterpart to the explicit multiple reflection series. This allows the application of an iterative solver like GMRES (Generalized Minimal Residual Method) [40, 41] to the derived underlying linear equation. GMRES converges reliably and quickly in contrast to the explicit multiple reflection series, often exhibiting poor convergence properties [24–26]. These poor convergence properties have been the main reason why the use of this procedure has not been common up to now—in contrast to the S-matrix algorithm.

Denominating the above algorithmic construction ‘S-vector algorithm’ by analogy with the ‘S-matrix algorithm’ emphasizes that the calculations take place only for a single incident vector of basis functions. The implicit variant of the S-vector algorithm promises to be often more efficient than the S-matrix algorithm of \(O(N^3)\)-complexity [21, 22, 27, 28], as well as more capable and flexible than the sole use of a fast integral equation-based method like GSM [12–16], which we demonstrate by examples in this manuscript.

This paper has the following structure:

- Chapter 2 presents the S-vector algorithm in explicit and implicit formulations, concentrating mainly on consecutive procedures in the framework of this publication. This chapter applies to 1D as well as 2D diffraction.

- Chapter 3 contains numerical investigations of the convergence behaviour as well as some potential applications. The examples will first refer to 1D grating diffraction for the sake of simplicity and intelligibility, with a final 2D application example.

- Chapter 4 concludes.

2. The scattering-vector algorithm

This chapter will start with a definition of the diffraction problem. Then, the S-vector algorithm is described—at first in parallel, then in detail in a more sophisticated sequential way.

2.1. Definition of the scattering problem

Figure 1 illustrates the geometry of the diffraction problem; it does this in 1D for sake of clarity, but the development below applies to both 1D and 2D structures. Light is incident solely from above onto a grating in the \(x\)-\(y\)-plane. The homogeneous medium on top of the grating is the medium of incidence and also called ‘cladding’ (cl.) The homogeneous medium below the grating is named ‘substrate’ (sub). The diffracting structure is divided into scattering layers along the \(z\)-direction.

Inside these layers—meaning for \(z_j < z < z_{j+1}\)—GSM [15, 16] is employed in the framework of the S-vector algorithm as one representative of a fast, rigorous scattering calculation method for reason of its availability. GSM works with the concept of a basis medium; in [15, 16], this is a homogeneous medium specified by the permittivity \(\varepsilon_r\) and the permeability \(\mu_r\) within layer \(j\). The basis functions (up and down functions, also called up and down amplitudes) inside
the \(j\)th layer are defined with respect to this basis medium. In the GSM case, the basis functions are plane waves and the eigenfunctions of the basis medium. Basis functions of such kind have already been applied in [3, 21, 26] between layers computed by the conventional differential method. The up and down amplitudes are denoted in this chapter by

\[
\begin{align*}
\mathbf{a}(z) : = & \left(\mathbf{a}_1(z), \mathbf{a}_j(z)\right)' \\
\Rightarrow & \left(\mathbf{a}_{1,z}(z), \mathbf{a}_{1,\text{b}}(z), \mathbf{a}_{j,z}(z), \mathbf{a}_{j,\text{b}}(z)\right)',
\end{align*}
\]

where the brackets ‘\(\cdot\)’ denote vectors of Fourier coefficients, the arrows ‘\(\uparrow\)’, ‘\(\downarrow\)’ refer to up- and downward propagation and the indices ‘\(\text{e}\)’, ‘\(\text{h}\)’ label TE- and TM-polarization. The definition of TE- and TM-polarization here coincides with [15, 16]. The entries of \(a(z)\) are the functions decomposing the field into upward propagating (and—where occurring—in \(+z\)-direction exponentially decreasing) harmonics and downward propagating (and—where occurring—in \(-z\)-direction exponentially decreasing) harmonics in the basis medium \((\varepsilon_j; \mu_j)\) in the layer \(j\). For \(z \leq 0\), the basis medium is the substrate and for \(z > h\) the cladding.

One basis medium \((\varepsilon_j; \mu_j)\) per layer \(j\) is defined in this paper. With a scattering-based formulation of the diffraction problem (see (2)–(5) below) being possible for non-magnetic as well as magnetic scatterers, the S-vector algorithm can be applied in both of these cases. The choice of this basis medium does not affect the correctness of the results, but has an influence on the number of GSM iterations required at the considered layer—see also chapter 3 and [43]. At each border between adjacent layers, the separation of the field into up and down functions as a requirement for the scattering-based calculations has to be enabled—similarly to [26]. This is accomplished by the theoretical insertion of an infinitesimally thin slice of the basis medium applied also inside the adjacent layer below and serves as an auxiliary device here. At the border between two basis media, the field amplitudes are scattered according to the Fresnel equations. The segmentation of the grating into scattering layers (scattering items) can be seen from figure 1: a diffracting region \(z_j < z < z_{j+1}\) together with the lower interface between two adjacent basis media will be condensed into one scattering item here. The uppermost scattering item additionally contains the jump between cladding and adjacent basis medium. The corresponding up and down amplitudes in two adjacent planes \(z_j, z_{j+1}\) can be related by scattering-matrix blocks (S-matrix blocks). \(S_{j+1,j}\) is the reflection matrix block and \(S_{j+1,j}^{(1)}\) the transmission matrix block for incidence from below; \(S_{j+1,j}^{(2)}\) is the reflection matrix block and \(S_{j+1,j}^{(2)}\) the transmission matrix block for incidence from above. The order of the superscripts of the S-matrix blocks is arbitrary here; for sake of uniformity, the smaller superscript will be on the right throughout.

Quantities referring to the \(n\)th diffraction order with respect to the \(x\)-direction and the \(n\)th diffraction order with respect to the \(y\)-direction are labelled by the indices \(n_x, n_y\), for which \(-N_x \leq n_x \leq N_x, -N_y \leq n_y \leq N_y\) hold, with \(N_x\) the maximum diffraction order considered in the \(x\)-direction and \(N_y\) the maximum diffraction order considered in the \(y\)-direction. In the case of 1D diffraction, \(N_y = 0\) and \(n_y = 0\) hold for all diffraction orders. The parameter \(N = (2N_x + 1) \cdot (2N_y + 1)\) denotes the number of diffraction orders in total. There are two \(k\)-vectors per diffraction order in the \(j\)th basis medium defined by \(\varepsilon_j, \mu_j\); one for upward propagation labelled by ‘\(\uparrow\)’ and one for downward propagation labelled by ‘\(\downarrow\)’. The modulus of the \(k\)-vector satisfies \(|k_{j,n_x,n_y}^{(\uparrow)}|^2 = |k_{j,n_x,n_y}^{(\downarrow)}|^2 = (k_j)^2 = \omega^2 \varepsilon_j \mu_j\), with \(\omega\) as angular frequency.

The expounded boundary value problem can be summarized by the following system of linear equations:

\[
\begin{align*}
\mathbf{a}_1^{(\uparrow)} = & \mathbf{S}_1^{(1\times-1)} \cdot \mathbf{a}_1^{(-1)} + \mathbf{S}_1^{(1\times-1)} \cdot \mathbf{a}_1^{(0)} \quad \quad \text{at the cladding} \\
\mathbf{a}_j^{(\uparrow)} = & \mathbf{S}_j^{(j\times-1)} \cdot \mathbf{a}_j^{(-1)} + \mathbf{S}_j^{(j\times-1)} \cdot \mathbf{a}_j^{(0)} \\
& \quad \quad \quad \quad \quad \quad \quad \text{for } 1 \leq j \leq s - 1 \\
\mathbf{a}_j^{(\downarrow)} = & \mathbf{S}_j^{(j+1\times)} \cdot \mathbf{a}_j^{(0)} + \mathbf{S}_j^{(j+1\times)} \cdot \mathbf{a}_j^{(j+1)} \\
& \quad \quad \quad \quad \quad \quad \quad \text{for } 1 \leq j \leq s - 1 \\
\mathbf{a}_0^{(\downarrow)} = & \mathbf{S}_0^{(1\times1)} \cdot \mathbf{a}_0^{(0)} + \mathbf{S}_0^{(1\times1)} \cdot \mathbf{a}_0^{(0)} \quad \quad \text{at the substrate},
\end{align*}
\]

where \(a_z^{(j)} := a_z(z_j)\) and \(a_{1,z}^{(j)} := a_z(z_{j+1})\). The incident light \(a_{1,z}^{(j)}\) and \(a_{1,z}^{(0)} = 0\) in (2)–(5) is given and \(a_{j,z}^{(j)}, a_{j,z}^{(j)}\) for \(1 \leq j \leq s - 1\) as well as \(a_{1,z}^{(\uparrow)}, a_{0,z}^{(\downarrow)}\) are the unknowns.

After sorting the summands (unknowns on the left, knowns on the right), one gets

\[
\begin{align*}
\mathbf{a}_j^{(\uparrow)} &= \mathbf{S}_j^{(j\times-1)} \cdot \mathbf{a}_j^{(-1)} + \mathbf{S}_j^{(j\times-1)} \cdot \mathbf{a}_j^{(0)} \quad \quad \text{at layer } s - 1 \\
\mathbf{a}_j^{(\downarrow)} &= \mathbf{S}_j^{(j+1\times)} \cdot \mathbf{a}_j^{(0)} + \mathbf{S}_j^{(j+1\times)} \cdot \mathbf{a}_j^{(j+1)} \\
& \quad \quad \quad \quad \quad \quad \quad \text{for } 1 \leq j \leq s - 1 \\
\mathbf{a}_0^{(\downarrow)} &= \mathbf{S}_0^{(1\times1)} \cdot \mathbf{a}_0^{(0)} + \mathbf{S}_0^{(1\times1)} \cdot \mathbf{a}_0^{(0)} \quad \quad \text{at the substrate},\end{align*}
\]
between a iterations would e.g. be required until J. Opt. 19

problem of incident harmonics. At a higher level, the boundary value

One way out involves the use of an iterative procedure eval-

In the case of a signif-

The stack of scattering layers defined above can be treated by the Bremmer series—or, rather, the multiple reflection series

First, some terminology is introduced. A ‘path’ is defined here by a sequence of scattering events of an incident harmonic at scattering layers until it exits the grating. A ‘scat-

In view of minimizing the memory demand, the focus is subsequently on the computation of a\(i\) merely for 1 \leq j \leq s – 1—meaning the up and down amplitudes between the layers. The reflection a\(i\) and the transmission a\(i\) follow later from a\(i\) and a\(i\) by (2) and (5).

Under certain conditions [23, 39, 46, 47] (see also next paragraph), the downwards propagating amplitudes a\(j\) at y\(j\) are the sum of all down amplitudes a\(j\) having undergone an even number 2\(i\) of reflections. Analogously, the upwards propagating amplitudes a\(i\) at y\(j\) are the sum of all up amplitudes a\(i\) having undergone an odd number 2\(i\) + 1 of reflections. This means

\[
a_{j}^{(i)} = \sum_{i=0}^{\infty} a_{j}^{(2i)}.
\]

\[
a_{i}^{(j)} = \sum_{i=0}^{\infty} a_{i}^{(2i+1)}.
\]
reflections. They can be calculated recursively:
\[
\mathbf{a}^{(j)}_{1,2i+1} = \sum_{l=1}^{s} P^{(j,l)}_{1} \cdot S^{(l-1)}_{1,1} \cdot \mathbf{a}^{(l)}_{1,2i+1},
\]
\[
\mathbf{a}^{(j)}_{1,2i} = \sum_{l=0}^{s} P^{(j,l)}_{1} \cdot S^{(l+1)}_{1,1} \cdot \mathbf{a}^{(l)}_{1,2i-1}
\]
with the matrices
\[
P^{(j,l)}_{1} := \begin{cases} 
S^{(l-1)}_{1,1} \ldots S^{(l+1)}_{1,1} & \text{for } l < j \\
1 & \text{for } l = j \\
0 & \text{for } l > j 
\end{cases}
\]
\[
P^{(j,l)}_{2} := \begin{cases} 
S^{(l-1)}_{1,1} \ldots S^{(l+1)}_{1,1} & \text{for } l > j \\
1 & \text{for } l = j \\
0 & \text{for } l < j 
\end{cases}
\]

‘P’ stands for ‘propagation’. Equations (12)–(18) specify the multiple reflection or Bremser series.

The series (12) converges if there is a norm \(\|\cdot\|\) and a real constant \(\zeta < 1\) such that \(\|a^{(2i+1)}_{1,2i+1}\| \leq \zeta \cdot \|a^{(2i)}_{1,2i}\|\) for all \(i \geq 0\) [46]; an analogous criterion holds for (13). According to our experience, this is usually fulfilled if at least a constant fraction of the light is coupled out of the grating in each upward or downward propagation step—if numerical errors do not play a role. In practice, the application of a sufficiently accurate and stable propagation method is essential when using the series (12)–(13); this is one motivation for the introduction of the respective implicit formulation (see next section).

Our implementation of the multiple reflection series-based propagation scheme utilized by the S-vector algorithm in the subsequent parts of this paper is described in appendix A. It has to be emphasized that we do not compute the S-matrix blocks in the framework of the S-vector algorithm in contrast to the multiple reflection series method in [24–26]. As well, we do not apply matrix-vector multiplications in case of the S-vector algorithm (see appendix A), which would cost \(O(N^2)\) operations. Instead, these matrix-vector products have to be understood as instructions to calculate the vector of transmitted or reflected amplitudes at a given layer from a vector of given incident amplitudes by means of a fast rigorous method—e.g. GSM [15, 16]. This is the novelty of the expounded explicit procedure at this point compared to the previous procedures described in [24–26].

In order to accomplish a more concise notation, alleviating our next steps, the following abbreviations are introduced. Skipping the superscript of the vector \(a\) means referencing all planes \(1 \leq j \leq s - 1\):
\[
a_1 := (a_1^{(1)}, \ldots, a_1^{(s-1)})^T, \quad a_j := (a_j^{(1)}, \ldots, a_j^{(s-1)})^T.
\]

The vector given at the beginning is then written as
\[
a_{\text{start}} := (0, 0, \ldots, S^{(s-1)}_{1,1} \cdot a^{(s-1)}_{1,1}),
\]
and the vector of the down amplitudes after step 0 (see figure 2 and appendix A) as
\[
a_{j,0} := (P^{(1,1)}_{1} \cdot a_{1,\text{start}}, \ldots, p^{(s-2,s-1)}_{1} \cdot a_{s-1,\text{start}}, a^{(1)}_{j,1})^T.
\]

The sub-matrix \(P^{(j)}_{1}\) propagates the down amplitudes downwards from plane \(j\) to plane \(j\).

The matrix \(P_{1}\) for downward propagation accumulates the intermediate (intm) down amplitudes \(a^{(j)}_{1,\text{intm},2}\)—where existing—from the observational planes \(l, 1 \leq l \leq s - 1\), and propagates them downwards to the observational planes \(j, 1 \leq j \leq s - 1\), so that
\[
a_{j,2i} = P_{1} \cdot a_{i,\text{intm},2i},
\]
where the components of \(a_{1,2i}\) are given by (16) and the components of \(a_{i,\text{intm},2}\) by \(a_{j,\text{intm},2j} = S^{(s-1)}_{1,1} \cdot a_{j,\text{start}}\) (see figure 2). The subscript ‘intm’ stands for ‘intermediate’; it labels quantities which are intermediate, not desired end results.

So, \(P_{1}\) consists of \((s - 1)^2\) sub-matrices of the dimension \((2N) \times (2N)\) with
\[
(P_{1})_{ij} := P^{(j,i)}_{1}
\]
in agreement with the arrangement of the sub-vectors in (19). For example,
\[
a_{i,0} = P_{1} \cdot a_{i,\text{start}}
\]
holds. The matrix \(P_{1}\) for ‘upward propagation’ is defined analogously:
\[
(P_{1})_{ij} := P^{(i,j)}_{1}
\]
so that
\[
a_{j,2i+1} = P_{1} \cdot a_{i,\text{intm},2i+1}.
\]
The application of which match the scheme of (A.12) explains the application of \( C_{11} \) and step 2 in appendix A (containing (A.9)–(A.12)) the application of \( C_{11} \).

Now, we can formulate the equations (12)–(16) in a way matching the application of \( C_0 = R_{o,0} \cdot P_o \):

\[
a_i = P_i \cdot \sum_{l=0}^{\infty} a_{i,\text{intm},2l}, \quad a_i = P_i \cdot \sum_{l=0}^{\infty} a_{i,\text{intm},2l+1}
\]

with the recursively defined summands

\[
a_{i,\text{intm},0} = a_{i,\text{start}}, \\
a_{i,\text{intm},2l+1} = C_{11} \cdot a_{i,\text{intm},2l}, \\
a_{i,\text{intm},2l} = C_{11} \cdot a_{i,\text{intm},2l-1}.
\]

The application of \( P_o \) has been postponed here to the respective subsequent step, in contrast to (12)–(16). Since the up and down amplitudes are mutually dependent, we eliminate \( a_{i,\text{intm},2l+1} \) from (37)–(39) in order to reduce the number of unknowns. This results in

\[
a_{i,\text{intm},0} = a_{i,\text{start}}, \\
a_{i,\text{intm},2l} = C_{11} \cdot a_{i,\text{intm},2l-2}.
\]

Hence, we obtain

\[
a_i = P_i \cdot \sum_{l=0}^{\infty} a_{i,\text{intm},2l} = P_i \cdot \sum_{l=0}^{\infty} C_{11} \cdot a_{i,\text{intm},2l-2} + a_{i,\text{intm},0} = P_i \cdot \sum_{l=0}^{\infty} C_{11} \cdot a_{i,\text{intm},2l-2} + a_{i,\text{intm},0}
\]

from (35)–(39) with

\[
C_{1 \downarrow \downarrow} = C_{11} \cdot C_{\downarrow \downarrow}.
\]
The upward propagating amplitudes are subsequently obtained by
\[ a_i = P_i \cdot a_{i,\text{start}} \quad (48) \]
and the reflected amplitudes finally by
\[ a_i^{(j)} = S_i^{(j)} \cdot a_i^{(j-1)} + S_i^{(j+1)} \cdot a_i^{(j-1)}. \quad (49) \]

2.3. The implicit formulation of the sequential S-vector algorithm

The calculation of \( \sum_{i=0}^{\infty} C_{[i]} \cdot a_{i,\text{start}} \) can be done in a more efficient and viable way than in (42) by switching to an implicit procedure:
\[ a_{i,\text{intm}} = \sum_{i=0}^{\infty} C_{[i]} \cdot a_{i,\text{start}} \quad (50) \]
\[ a_i = (1 - C_{[i]})^{-1} \cdot a_{i,\text{start}} \quad (51) \]
\[ (1 - C_{[i]}) \cdot a_{i,\text{intm}} = a_{i,\text{start}}. \quad (52) \]

Since the application of a direct solver, computing the inverse of an explicitly specified matrix at a clearly higher than linear complexity, would not match the concept of vector-based fast calculations here, (52) is solved iteratively. Any iterative solver can be tried. Since GMRES [40, 41] is an already-established solver for the solution of scattering problems in the case of GSM [15, 16], it will be applied in the framework of this paper. Compared to other iterative solvers, GMRES has a reliable, optimal convergence—although this is at the expense of a memory demand proportional to the number of iterations [40, 41]. Once \( a_{i,\text{intm}} \) has been computed, all other quantities can be computed in the same way, as described in the last paragraph of the preceding section by (45)–(49).

Concluding, the implicit formulation (52) is also very memory sparing: instead of both up and down amplitudes at the layer boundaries, as in e.g. (6)–(11), only half of them (merely down amplitudes) are stored during the iterative solution. A summary of the algorithmic steps required by the approach (52) is listed in appendix B. For the sake of completeness, we mention that there is an equivalent to (52) when choosing \( a_{i,\text{intm}} = C_{[i]} \cdot a_{i,\text{intm}} \) instead of \( a_{i,\text{intm}} \) as vector of unknowns, which is obtained from (52) by multiplication with \( C_{[i]} \). The completion, which is required for incidence from both sides, is obtained by the application of the superposition principle.

To our knowledge, the reformulation of the Bremmer (multiple reflection) series method in an implicit way, and its solution by an iterative solver, is new.

We call the solution of (52) by an iterative solver ‘implicit variant of the sequential S-vector algorithm’ in this paper—provided that the calculations at the single layers are conducted by a method of linear complexity in time and memory; it is an alternative to the solution of (6)–(11) by an iterative solver, which we call ‘implicit variant of the parallel S-vector algorithm’.

In what follows, we give the merits of the implicit formulations.

The implicit formulation (52) can also be derived directly from (2)–(5) by inserting \( a_i^{(j)} \) sequentially for decreasing \( j \) into the other equations and then inserting \( a_i^{(j)} \) sequentially for increasing \( j \) into the other equations; in contrast to the multiple reflection series, this does not require the expandability of an inverse matrix into a series and the conditions connected with this.

If GMRES is taken as iterative solver, the convergence is ensured as long as the matrix on the left side of (52) is invertible (and the computational accuracy is sufficient), since GMRES yields the exact solution of a linear system of equations of full rank, if the number of iterations is equal to the number of unknowns and rounding errors are absent [40, 41]. Fortunately, the problem (52) is usually well-conditioned due to its scattering-based formulation. Hence, the number of iterations required for a certain accuracy is normally much lower than the number of unknowns, and rounding errors play a minor role, so that even single precision computations are sufficient. In contrast to this, obtaining results of a desired accuracy is not guaranteed when employing the explicit procedure (44) and truncating the series in practice.

A more subtle advantage of the implicit formulation (52) compared to the explicit procedure (44) is a higher accuracy of the results and, especially, a better stability of the method. The implicit formulation (52) requires propagation of the light only once between cladding, substrate and back in one iteration. So, the propagation is conducted only over a very limited distance. In contrast to this, the explicit procedure (44) requires multiple-fold propagation between substrate and cladding—meaning propagation over a long distance. Since the propagation or integration errors at the layers accumulate, this can lead to a large total error. A very accurate and stable propagation procedure is therefore crucial in this case. We think that the divergence of the multiple reflection series in [26] is caused by such integration errors (although we have not tried to reproduce this computation). We are convinced that this divergence results from a combination of insufficient stability of the applied numerical integration method at the layers with the multiple reflection series (44) as an explicit procedure that leads to a strong error accumulation. Having always combined the S-vector algorithm with GSM [15, 16], such divergences are not part of our wealth of experience.

3. Numerical examples on the basis of the sequential S-vector algorithm and discussion

This chapter investigates the convergence of the sequential S-vector algorithm on the basis of numerical examples (first examples for the parallel S-vector algorithm are given in [44]). Subsequently, two examples for the beneficial use of the superior, reliably converging implicit formulations are given. GSM [15, 16] as one representative for a fast, rigorous method is used for the computations at the single scattering layers.

The examples presented here refer to glass gratings. We have also applied the combination of the S-vector algorithm
with GSM to gratings of higher refractive index contrast, such as gratings consisting of bumps of chromium and grooves of air, and observed convergence. Yet, GSM is less efficient and therefore less suited for gratings of high refractive index contrast and metal gratings than for gratings of moderate refractive index contrast, since GSM needs a higher number of iterations in such cases. This motivates work on efficient methods for gratings of high refractive index contrast.

With the presented examples referring to non-magnetic scatterers, we choose $\mu_j = \mu_{\text{vac}}$ (permeability of the vacuum) there.

The computation results for diffraction at 1D gratings have here been checked by comparison with the differential method [3, 4] and the ones for diffraction at a 2D grating (beam splitter) by RCWA. For the implicit formulation (52), we have also made comparisons with the sole application of GSM to gratings (without the S-vector algorithm). Here, the differences in the outgoing diffracted complex amplitudes between use of and no use of the S-vector algorithm are roughly of the order of the residuals with which the equations are solved iteratively. Consequently, the use of the implicit formulations does not falsify results, and so the implicit formulations are correct.

All computations here are conducted using double precision variables. However, all inversions are performed merely with single precision:

- As termination criterion for the explicit variant (44) of the S-vector algorithm we chose
  \[ \| \mathbf{C}_{\text{diff}}^{\text{new}} \cdot \mathbf{a}_i^{\text{start}} \|_2 \leq 10^{-7} \cdot \| \mathbf{a}_i^{\text{old}} \|_2 = 10^{-7}, \]  (53)
  \[ |i|_2 \text{ denoting the 2-norm. The vector } \mathbf{a}_i^{\text{old}} \text{ of light incident onto the grating has the modulus 1.} \]
- In the case of the variant (52) referred to as the ‘implicit’ one, we chose as termination criterion a GMRES residual equal to or smaller than $10^{-7}$:
  \[ \text{residual}_{\text{GMRES}}(\mathbf{a}_i^{\text{start}}) \leq 10^{-7}. \]  (54)

A residual equal to or smaller than $10^{-7}$ is also applied to GSM as termination criterion; $10^{-7}$ is roughly in the order of float precision. The specified termination criteria are sufficient here due to the stability of the employed methods.

All inversions here are executed by GMRES as one representative of an iterative solver. The zero vector is taken as initial guess in case of all presented examples.

### 3.1. Convergence behaviour

This subsection investigates the convergence behaviour of the sequential S-vector algorithm under various conditions, and especially compares the explicit formulation (44) with the implicit one (52). The test cases are depicted in figure 3, and are listed below.

(i) The first test case concerns almost perpendicular incidence onto a 1D Ronchi phase grating, which is frequently applied in interferometric setups. The vacuum wavelength is $\lambda_{\text{vac}} = 632.8 \text{ nm}$; in the case of fused silica as a grating material, this requires a height of $h = 692.2 \text{ nm}$ of the modulated region. The angle of incidence is $0.7^\circ$ here, and would avoid an undesired back reflection in practical applications. The modulated region is divided into $s = 2$ layers of thickness $h/2 = 346.1 \text{ nm}$. Between these layers, there are infinitesimally thin slices of a basis medium (also called background medium) as auxiliary device—see section 2.1 and [26]. The refractive index of the basis medium above a layer $j$ is $n_j = (1 + 1.4571)/2 = 1.2286$ (the permeability is $\mu_j = \mu_{\text{vac}}$). Generally, a good choice for the refractive index of the basis medium is the average refractive index.

(ii) The second test case differs from the first one merely in the number of layers the grating is divided into. Here, $s = 8$ layers of equal thickness $h/8 = 86.525 \text{ nm}$ are taken. They are separated by the above mentioned infinitesimally thin slices of the basis medium.

(iii) The third test case differs from the first one in the thickness of the modulated region, which is $h = 4 \cdot 692.2 = 2768.8 \text{ nm}$ here, and in the number of layers, which is $s = 8$. Their thickness is
Table 1. Diffraction efficiencies as a function of the diffraction order in reflection and transmission for TE-polarization in case of the grating and numerical parameters depicted in figure 3 for $h = 692.2$ nm. The results have been obtained by a combination of the S-vector algorithm with GSM at the single layers, and checked by the differential method [3, 4].

| Diffraction order | -2   | -1   | 0    | 1    | 2    |
|-------------------|------|------|------|------|------|
| Reflection TE     | 0.0020 | 0.0010 | 0.0812 | 0.0017 | 0.0009 |
| Transmission TE   | 0.3971 | 0.1178 | 0.3984 |—     |—     |

$h/8 = 2768.8/8$ nm = 346.1 nm each. They are separated by the above mentioned infinitesimally thin slices of the basis medium.

The diffraction efficiencies for the test cases are listed in tables 1 and 2.

We consider these three test cases in order to investigate the convergence of the S-vector algorithm variants under various conditions: the first case represents the scenario of a low number of rather strongly scattering layers, the second case the scenario of a large number of thin, rather weakly scattering layers and the third case the scenario of a large number of thick, rather strongly scattering layers. A comparison between test cases 1 and 2 allows practical recommendations whether to divide a structure of given thickness into many thin (weakly scattering) layers or few thick (strongly scattering) ones. Test case 3 is listed as an extreme case where two features slowing down the convergence are combined: a large number of layers as well as strong reflections at these layers due to their large thickness. Here, the explicit variant (44) arrives at its limits in practice.

All GSM computations applied to entire gratings as well as single layers have been conducted using a 4th-order integration technique based on midpoint rule and extrapolation (step sizes $\Delta z = 692.2$ nm /64 = 10.82 nm and $\Delta z = 692.2$ nm /96 = 7.210 nm or respectively $\Delta z = 2768.8$ nm /256 = 10.82 nm and $\Delta z = 2768.8$ nm /384 = 7.210 nm). This yields convergence order 4 with respect to $\Delta z$ instead of 2 for sole use of the midpoint rule—similarly to chapter 4 (‘numerical examples and discussion’) in [43].

The convergence behaviour of the two variants (44), (52) is presented in figures 4–6, each for all three test cases and TE-polarization. (We made the investigations for TM-polarization for the considered examples as well; here, the convergence for TM-polarization turned out to be slightly better, which we attribute to smaller reflection coefficients at the layer boundaries for TM-polarization.)

In the case of the explicit variant (44), the residual defined in (53) is plotted as a function of the truncation number $i_{\text{max}}$. In the case of the implicit variant (52), the GMRES residual defined in (54) is plotted as a function of the iteration number. Regarding the explicit variant (44) and the implicit one (52), an increment of one on the $x$-axis means another single-fold application of $C_{[\text{III}]}$ or $1 - C_{[\text{III}]}$ respectively.

Table 2. Diffraction efficiencies as a function of the diffraction order in reflection and transmission for TE-polarization in case of the grating and numerical parameters depicted in figure 3 for $h = 2768.8$ nm. The results have been obtained by a combination of the S-vector algorithm with GSM at the single layers, and checked by the differential method [3, 4].

| Diffraction order | -2   | -1   | 0    | 1    | 2    |
|-------------------|------|------|------|------|------|
| Reflection TE     | 0.0051 | 0.0038 | 0.0335 | 0.0039 | 0.0031 |
| Transmission TE   | 0.0044 | 0.9417 | 0.0046 |—     |—     |

Figure 4. Convergence of the explicit (44) and the implicit (52) formulation of the S-vector algorithm for test case 1 (2 layers of 692.2/2 nm = 346.1 nm thickness each) and TE-polarization. For comparison: sole application of GSM to the entire grating takes 23 iterations here in case of 64, as well as 96 sampling points in the $z$-direction.

Figure 5. Convergence of the explicit (44) and the implicit (52) formulation of the S-vector algorithm for test case 2 (8 layers of 692.2/8 nm = 86.525 nm thickness each) and TE-polarization. For comparison: sole application of GSM to the entire grating takes 23 iterations here in case of 64, as well as 96 sampling points in the $z$-direction.
A comparison between the explicit (44) and the implicit variant (52), both involving the same matrix \( C_{111} \), shows a much better convergence behaviour of the implicit technique, particularly in test case 3 (figure 6). The implicit procedure converges super-linearly, whereas the explicit procedure converges maximally linearly, and exhibits a trend towards stagnation in some places. All in all, the benefit of the implicit variant can be seen, as expected.

Next, the division of a grating of given height \( h \) into a small number and a large number \( s \) of layers are compared with each other. This refers to the comparison of test cases 1 with 2 and thus of figure 4 with 5. The comparison indicates that the division of a grating of given height \( h \) into a large number \( s \) of layers results in a deteriorated convergence behaviour of all considered variants (44), (52) compared to a division into a small number \( s \) of layers. The physical interpretation of this is that the waves are reflected multiple times at the layers before leaving the grating, particularly in the case of a high number of scattering layers. This motivates the choice of a small \( s \geq 2 \) (meaning possibly thick layers) and—where required—the use of a recursive procedure (division of single possibly thick layers in figure 1 into sub-layers to be computed separately and to be combined by the S-vector algorithm to layers).

Finally, the effect of the grating thickness \( h \) on the convergence is observed. This refers to the comparison of test cases 1 and 3. The scattering layers are identical in both cases, but their number \( s \) is different and thus also the height of the entire grating. The comparison of figure 4 with 6 shows that a larger number of scattering layers (or rather a larger height \( h \)) results in slower convergence. The physical interpretation here is similar to the one given in the preceding paragraph. So, the efficiency of the applied combination of GSM with the S-vector algorithm has a tendency to decrease with increasing grating thickness (the sole application of GSM has such a tendency here, too). Yet, gratings of a thickness in the order of a few wavelengths are rather rare in practice—apart from waveguides (or rather fibers), where modal methods are a good choice.

All in all, the implicit variant of the S-vector algorithm converges reliably and very well.

### 3.2. The S-vector algorithm as a means of enhancing the flexibility of integral equation-based methods

This section outlines some potential applications of the S-vector algorithm beyond a sole reduction of the memory demand. A feature of the S-vector algorithm is its flexibility and generality; the requirements for its application are the presence of two or more scattering items and the representation of the field at or between these items in terms of forward and backward functions—e.g. plane waves, modal basis functions, spherical waves etc. Hence, it is particularly suited for the combination of scattering layers of different material and shape or of a layer with scattering particles on top of it.

In the framework of this paper, GSM [15, 16] based on an integral equation is applied for the computations at the layers. Methods based on integral equations have mainly been connected with the application of equidistant sampling [12–16] for reasons of simplicity and efficiency until now. The radiation operators in [12–16] have a Toeplitz structure, then. Here, the S-vector algorithm can be used for an adaptation of numerical parameters like the integration step size—e.g. a smaller step size for metallic layers than for dielectric ones. It can be used, as well, for precise adaption of the integration mesh to the thicknesses of the layers. Besides, it allows the individual choice of the basis medium for each layer.

The following simulation example refers to a 1D 4-level phase grating with layers of different thicknesses due to fabrication errors (figure 7). Application of equidistant meshing between substrate and cladding with a constant step size of a few nanometers seems not advisable here, since the heights of the stairs differ from each other. This motivates an individual meshing at each layer, and a combination of the layers by the S-vector algorithm. Moreover, all GSM computations at each layer are conducted twice here with different step sizes \( \Delta z \) (see figure 7 on the left, as well as the corresponding caption) and the results are extrapolated as in the previous subsection for an accuracy enhancement. The basis medium is specified for each layer separately (see refractive indices \( n_b \) in figure 7). The segmentation of the grating into scattering items can be seen from figures 1 and 7: the uppermost scattering layer is comprised of the transition between cladding and basis medium of refractive index \( n_{b,3} \), the uppermost binary profile and the transition from the basis medium of refractive index \( n_{b,3} \) to \( n_{b,2} \). The intermediate scattering layer consists of the intermediate modulated layer and the transition from basis medium \( n_{b,2} \) to \( n_{b,1} \). The lowermost scattering layer is comprised of the lowermost binary
onto a 4-level 1D phase grating of fused silica

Figure 7. Test case: non-conical, oblique incidence at an angle of \( \theta = 40^\circ \) onto a 4-level 1D phase grating of fused silica (\( n_{\text{sub}} = 1.4571 \)) and air (\( n_{\text{air}} = 1 \)). The maximum diffraction order is \( N = 150 \), the length of period is \( p = 1000 \) nm and the height is \( h = 1045.2 \) nm. TE- and TM-polarization are considered in separate tests. The basis medium at the lowermost stair has the refractive index \( n_{b,1} \), the one at the intermediate stair \( n_{b,2} \) and the one at the uppermost stair \( n_{b,3} \). The GSM computations at the single layers are combined with an extrapolation technique for an accuracy enhancement [45] as in the previous subsection. This requires numeric integration with 2 different step sizes in the z-direction each time a vector of harmonics is incident onto a layer. The respective numbers of z-sampling points are listed on the left of this figure.

profile and the transition from the refractive index \( n_{b,i} \) to the substrate.

The number of S-vector iterations is merely seven for both polarizations, using the implicit formulation (52). The diffraction efficiencies for the grating depicted in figure 7 are listed in table 3.

The main purpose of the preceding sections was the investigation and explanation of basic features and aspects of the S-vector algorithm; this took place on the basis of 1D examples. In the next section, we turn to an application example in 2D in combination with computations on a graphic card. The method applied at the layers remains GSM—but now in the formulation for 2D diffraction [16].

3.3. The S-vector algorithm as memory sparing technique facilitating GPU computations of 2D grating diffraction

In recent years, graphic card (GPU) computations have become an appealing alternative to central processor (CPU) computations. Yet, a critical point even nowadays is memory limitation on GPUs (4 GB in case of the GPU used in this paper). A higher memory demand necessitates the use of the RAM on the motherboard, and therefore time-consuming data exchanges between GPU and motherboard. In the S-vector implementation utilized in this section, we conduct the computationally heavy GSM computations at the single layers of a 2D grating on a GPU and use the RAM only when the GPU memory becomes insufficient. These are the places where the RAM usage increases rapidly in tables 4 and 5. The S-vectors are always stored in the RAM, hardly occupying any memory (see table 5 for 51^2 and 101^2 diffraction orders). The implicit formulation is applied and tested.

The subsequent example shows that the memory demand of GSM can be strongly reduced by the S-vector algorithm. Normal incidence of light of the wavelength \( \lambda_{\text{vac}} = 1000 \) nm onto a beam splitter of \( 2 \times 2 \) pixels is considered (see figure 8). The incident electric field at \( (0; 0; h) \) is parallel to the y-axis. The pixels as well as the substrate consist of glass of \( n_{\text{sub}} = 1.5 \). The periods are \( p_y = p_x = 7.5294 \) \( \mu \text{m} \). The height \( h = 1000 \) nm would effect a phase shift of \( \pi \) in the ray optical limit at the glass pixels, and thus a diffraction efficiency of 0 in the zeroth order. The rigorously computed diffraction efficiency in the zeroth order is 0.02016; the efficiency in the orders \((-1, -1), (-1, 1), (1, -1), (1, 1)\) is 0.15655 each, summarizing TE- and TM-portions.

We conduct the numerical integration with 120 and with 150 sampling points equally distributed over the height \( h \) and extrapolate for the limit of an infinite number of sampling points. Subsequently, we compare the sole use of GSM (see table 4) with the S-vector algorithm combined with GSM (see table 5) for an increasing number of diffraction orders. The number of S-vector layers is three here, each one containing 150/3 = 50 and 120/3 = 40 sampling points along the z-direction.

The first line of table 4 shows a very high number of GMRES iterations (above 300) required by GSM. In combination with a high number of integration sampling points (150 and 120), the memory demand of GSM is very high. So, the RAM is utilized to some extent already for 51^2 diffraction orders (per polarization). Only the circumstance that our workstation has 64 GB RAM allows the computation of 151^2 diffraction orders; 201^2 orders cannot be computed any more. The 4th line of table 4 demonstrates that the computation time increases faster than linearly with the number of diffraction orders, due to an increasing use of the RAM and time-consuming data exchanges between GPU and RAM. This occurs already from the beginning of our test with 51^2 diffraction orders. The first line of the second table shows a much lower number of iterations required by GSM at a single layer (not more than 100 iterations, since the layers have only a thickness of \( h/3 \)). In combination with a much lower number of integration sampling points (50 and 40), the memory demand is lower here by about a factor of nine, compared to the sole use of GSM. Hence, the RAM utilization is much lower. No RAM is used for GSM computations at single layers for up to 151^2 diffraction orders. The column belonging to 101^2 diffraction orders shows that the RAM occupancy by the S-vector algorithm is negligible here; the respective down amplitudes are considered merely in the two planes between the three layers in figure 8. A computation with 501^2 orders (per polarization) is possible. This is a considerable step towards the computation of large 2D objects. The deviation of the sum of efficiencies from 1 is merely \( 7 \times 10^{-8} \) here, which is in the order of the residuals of \( 10^{-6} \) with which all the systems of equations are solved here. This is further evidence for the correctness of the methods applied here.
Table 3. Diffraction efficiencies as a function of the diffraction order in reflection and transmission for TE- and TM-polarization in case of the grating and the numerical parameters depicted in figure 7. The results have been obtained by a combination of the S-vector algorithm with GSM at the single layers and been checked by the differential method [3, 4].

| Diffraction order | -3 | -2 | -1  | 0   | 1   |
|-------------------|----|----|-----|-----|-----|
| Reflection TE     | —  | 0.0003 | 0.0002 | 0.0301 | —  |
| Transmission TE   | 0.0234 | 0.0323 | 0.6930 | 0.1635 | 0.0571 |
| Reflection TM     | —  | 0.0006 | 0.0003 | 0.0094 | —  |
| Transmission TM   | 0.0185 | 0.0278 | 0.6760 | 0.2256 | 0.0419 |

Table 4. Performance of GSM. First line: number of diffraction orders $(2N_x + 1) \cdot (2N_y + 1)$; $N_x = N_y$ is chosen. Second line: highest number of iterations required by GSM. Third line: peak RAM usage in GB. Forth line: total computation time in hours.

| $(2N_x + 1) \cdot (2N_y + 1)$ | $51^2$ | $101^2$ | $151^2$ | $201^2$ |
|-------------------------------|--------|--------|--------|--------|
| Highest no. of GSM iterations | 306    | 311    | 310    | —      |
| Peak RAM usage/GB             | 1.8    | 18.7   | 47.0   | —      |
| $t/h$                         | 0.043  | 0.56   | 2.34   | —      |

For comparison, we have conducted an RCWA-computation of the $2 \times 2$-beam splitter with $51^2$ diffraction orders on a CPU (not using any GPU), which may be considered as a standard. The memory demand has been 13.7 GB here and the computation time 1.8 hours. A computation with $101^2$ diffraction orders has not even been possible on our workstation, due to a lack of memory.

4. Conclusion

A novel scattering formalism called ‘S-vector algorithm’ has been presented in its consecutive variants (explicit and implicit). The new formalism can be considered as a fast, iterative counterpart to the S-matrix algorithm of $O(N^3)$-complexity, and as a complement to existing fast scattering formalisms based on integral equations.

The new formalism differs from the multiple reflection series [24–26] by two significant improvements: firstly, the vectors of up and down functions are diffracted at the scattering items by a fast method (here: GSM) instead of a slow method computing entire S-matrix blocks with $O(N^3)$-complexity as a preliminary measure. This results in $O(N \ln(N))$-complexity of the method overall. Secondly, the implicit counterpart to the multiple reflection series has been derived. This allows the application of an iterative solver like GMRES, and ensures convergence (provided that the method for the scattering computations at the single scattering items converges). The very good convergence, particularly in the case of a low number of layers, also suggests iterative solvers different from GMRES here.

The computational load of the entire calculation procedure is proportional to the number of S-vector iterations, as well as to the number of iterations at the single layers; it is dominated by the complexity of the method employed at the single scattering objects, which should ideally be linear with a small constant. The memory demand of the S-vector algorithm is linear with respect to the number of considered basis functions, and proportional to the number of iterations of which the result is stored by the iterative solver.

Some examples for the use of the presented method have been given here: the S-vector algorithm allows a drastic reduction of the memory consumption (at the expense of a higher number of operation counts) by division of gratings into layers to be computed separately. This makes it especially appealing for graphic card computations; in the case of computations of large 2D gratings (e.g. with $501^2$ harmonics) on graphic cards, the memory demand is indeed the limiting criterion. Another application example is non-equidistant meshing in combination with higher-order integration. The expounded method also allows the combination of scattering objects of different shape (e.g. spherical ones), size or material, suggesting utilization of different methods or the adaptation of numerical parameters to these objects. Such parameters are e.g. integration step size and basis medium in the case of use of GSM.

Summarizing, the S-vector algorithm extends the possibilities of fast grating computation by enabling more flexibility. We think that the proposed method will gradually become more important with an increasing variety of fast methods available, and a growing demand for scattering computations at complex-shaped, large scattering objects.

Acknowledgments

The first author is grateful to Prof N Lindlein, supervisor of his PhD thesis at the University of Erlangen-Nuremberg, and his group for the constant support as well as the funding of his PhD thesis. The commitment of Prof Emeritus O Parriaux, Saint-Etienne University, and Prof Gérard Granet, Clermont-Ferrand University, in reviewing the manuscript is acknowledged. This work was supported by the LABEX MANUTECH-SISE (ANR-10-LABX-0075) of Université de Lyon, within the program ‘Investissements d’Avenir’ (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR).
Table 5. Performance of GSM combined with the S-vector algorithm. First line: number of diffraction orders \((2N_i + 1) \cdot (2N_f + 1); N_i = N_f\) is chosen. Second line: highest number of iterations required by GSM at any single layer. Third line: peak RAM usage in GB. Fourth line: no. of S-vector iterations. Fifth line: total computation time in hours.

|                        | \((2N_i + 1) \cdot (2N_f + 1)\) | \(51^2\) | \(101^2\) | \(151^2\) | \(201^2\) | \(\cdots\) | \(501^2\) |
|------------------------|---------------------------------|---------|---------|---------|---------|---------|---------|
| Highest no. of GSM iterations | 98                              | 99      | 99      | 100     | 100     | 100     | 100     |
| Peak RAM usage/GB      | 0.2                             | 0.2     | 1.9     | 6.4     | 60.0    |         |         |
| no. of S-Vector iterations | 47                              | 47      | 47      | 47      | 47      |         |         |
| \(t/h\)                | 0.24                           | 0.84    | 2.24    | 7.37    | 46.65   |         |         |

For all \(j\) from \(j = s\) to \(j = 1\):
- if \(j > 1\): \(a_{i,0}^{(j-1)} = S_{j,0}^{(i-1)} \cdot a_{i,0}^{(j)}\), where \(a_{i,0}^{(0)} = a_i\); \(\text{(A.1)}\)
  - add \(a_{i,0}^{(j-1)}\) to \(a_i^{(j-1)}\). \(\text{(A.2)}\)
- if \(j < s\): \(a_{i,1}^{(j+1)} = S_{j+1,1}^{(j)} \cdot a_{i,1}^{(j)}\); \(\text{(A.3)}\)
  - buffer \(a_{i,1}^{(j)}\) for step 1. \(\text{(A.4)}\)

**Step 1:** Propagation of the single-fold reflected light upwards to the cladding (see figure 2):

Start with \(a_{i,1}^{(1)} = 0\). All other \(a_{i,j}^{(1)}\) will follow from the algorithm due to (A.8) and increasing \(j\).

For all \(j\) from \(j = 1\) to \(j = s - 1\):
- add buffered \(a_{i,1}^{(j)}\) to \(a_{i,2}^{(j)}\). \(\text{(A.5)}\)
- add \(a_{i,0}^{(j)}\) to \(a_{i,1}^{(j)}\). \(\text{(A.6)}\)
- \(a_{i,2}^{(j)} = S_{j+1,1}^{(j)} \cdot a_{i,2}^{(j)}\); buffer \(a_{i,2}^{(j)}\) for step 2. \(\text{(A.7)}\)
- if \(j < s - 1\): \(a_{i,1}^{(j+1)} = S_{j+1,1}^{(j+1)} \cdot a_{i,1}^{(j+1)}\) \(\text{(A.8)}\)

**Step 2:** Propagation of the double-fold reflected amplitudes downwards to the substrate (see figure 2):

Start with \(a_{i,2}^{(1)} = 0\). All other \(a_{i,j}^{(2)}\) will follow from the algorithm due to (A.12) and decreasing \(j\).

For all \(j\) from \(j = s - 1\) to \(j = 1\):
- add buffered \(a_{i,2}^{(j)}\) to \(a_{i,3}^{(j)}\). \(\text{(A.9)}\)
- add \(a_{i,0}^{(j)}\) to \(a_{i,1}^{(j)}\). \(\text{(A.10)}\)
- \(a_{i,3}^{(j)} = S_{j+1,2}^{(j)} \cdot a_{i,3}^{(j)}\); buffer \(a_{i,3}^{(j)}\) for step 3. \(\text{(A.11)}\)
- if \(j > 1\): \(a_{i,2}^{(j-1)} = S_{j+1,2}^{(j-1)} \cdot a_{i,2}^{(j-1)}\) \(\text{(A.12)}\)

**Subsequent steps:**

Multiple application of steps 1 and 2 sequentially, until the modulus of \(a_{i,j}^{(2j+1)}\) is negligibly small \(\forall j\) fulfilling \(1 \leq j \leq s - 1\) (or until this holds for \(a_{i,j}^{(2)} \forall j\) fulfilling \(1 \leq j \leq s - 1\)).

Figure 8. 2D beam splitter of \(2 \times 2\) pixels of equal size under normal incidence. The incident electric field \(E^{(inc)}\) is parallel to y-direction at \((0; 0; h)\). Pixels and substrate are of the same material (glass of refractive index 1.5). The height is \(h = 1000\) nm. The modulated region is divided into three S-vector layers of equal thickness. The periods are \(p_x = p_y = p = 7.5294\) \(\mu m\). We choose \(N_i = N_f\) here for all computations.

Appendix A. Implementation of the applied stepwise propagation procedure

This appendix lists our implementation of the stepwise downward and upward propagation of the explicit variant of the sequential S-vector algorithm, described in section 2.2 and depicted in figure 2; the algorithm calculates the up and down amplitudes \(a_{i,j}^{(1)}, a_{i,j}^{(2)}, 1 \leq j \leq s - 1\) between the layers.

Fast methods usually yield the reflection at a layer together with the transmission. Hence, it is efficient to buffer the reflected amplitudes obtained during the application of \(P_2\) so that they are available without repeated computation in the next step of the recursive algorithm. The respective procedure is expounded below. The subscript ‘intm’ will stand for ‘intermediate’; it will label quantities which are intermediate, not desired end results.

_Initialization:_

Start with \(a_{i,1}^{(0)} = 0, a_{i,2}^{(0)} = 0\) for \(j = 1, \cdots, s - 1\).

_Step 0:_ Propagation of the incident amplitudes \(a_{i,j}^{(s)}\) downward to the substrate (see figure 2):

For all \(j\) from \(j = s\) to \(j = 1\):
- if \(j > 1\): \(a_{i,0}^{(j-1)} = S_{j,0}^{(i-1)} \cdot a_{i,0}^{(j)}\), where \(a_{i,0}^{(0)} = a_i\); \(\text{(A.1)}\)
  - add \(a_{i,0}^{(j-1)}\) to \(a_i^{(j-1)}\). \(\text{(A.2)}\)
- if \(j < s\): \(a_{i,1}^{(j+1)} = S_{j+1,1}^{(j)} \cdot a_{i,1}^{(j)}\); \(\text{(A.3)}\)
  - buffer \(a_{i,1}^{(j)}\) for step 1. \(\text{(A.4)}\)

_Step 1:_ Propagation of the single-fold reflected light upwards to the cladding (see figure 2):

Start with \(a_{i,1}^{(1)} = 0\). All other \(a_{i,j}^{(1)}\) will follow from the algorithm due to (A.8) and increasing \(j\).

For all \(j\) from \(j = 1\) to \(j = s - 1\):
- add buffered \(a_{i,1}^{(j)}\) to \(a_{i,2}^{(j)}\). \(\text{(A.5)}\)
- add \(a_{i,0}^{(j)}\) to \(a_{i,1}^{(j)}\). \(\text{(A.6)}\)
- \(a_{i,2}^{(j)} = S_{j+1,1}^{(j)} \cdot a_{i,2}^{(j)}\); buffer \(a_{i,2}^{(j)}\) for step 2. \(\text{(A.7)}\)
- if \(j < s - 1\): \(a_{i,1}^{(j+1)} = S_{j+1,1}^{(j+1)} \cdot a_{i,1}^{(j+1)}\) \(\text{(A.8)}\)

_Step 2:_ Propagation of the double-fold reflected amplitudes downwards to the substrate (see figure 2):

Start with \(a_{i,2}^{(1)} = 0\). All other \(a_{i,j}^{(2)}\) will follow from the algorithm due to (A.12) and decreasing \(j\).

For all \(j\) from \(j = s - 1\) to \(j = 1\):
- add buffered \(a_{i,2}^{(j)}\) to \(a_{i,3}^{(j)}\). \(\text{(A.9)}\)
- add \(a_{i,0}^{(j)}\) to \(a_{i,1}^{(j)}\). \(\text{(A.10)}\)
- \(a_{i,3}^{(j)} = S_{j+1,2}^{(j)} \cdot a_{i,3}^{(j)}\); buffer \(a_{i,3}^{(j)}\) for step 3. \(\text{(A.11)}\)
- if \(j > 1\): \(a_{i,2}^{(j-1)} = S_{j+1,2}^{(j-1)} \cdot a_{i,2}^{(j-1)}\) \(\text{(A.12)}\)
The transmission into the substrate follows subsequently by $\mathbf{a}_{n}^{(l)} = S_{l}^{(l,0)} \cdot \mathbf{a}_{l}^{(0)}$ and the reflection into the cladding by $\mathbf{a}_{n}^{(r)} = S_{l}^{(l,r-1)} \cdot \mathbf{a}_{l}^{(r-1)} + S_{l}^{(l,r-1)} \cdot \mathbf{a}_{l}^{(r)}$.

Appendix B. Implementation of the implicit variant of the sequential S-vector algorithm

Regarding the implicit variant defined by (52), (43) and (20), step 2 in appendix A without the item (A.10) describes the application of $C_{l}$ and step 1 without the item (A.6) the application of $C_{l}$. Setting up the required vector $\mathbf{a}_{\text{start}}$ defined in (20) merely requires a single transmission calculation of the incident down amplitudes $\mathbf{a}_{n}^{(l)}$ at the uppermost layer $s$ by GSM. With these implementations, an iterative solver such as GMRES [40, 41] can be applied to (52) with $\mathbf{0}$ as initial guess for the unknown $\mathbf{a}_{l}^{(\text{int})}$.

The remaining quantities finally follow from $\mathbf{a}_{l}^{(\text{int})}$ by (45)–(49).

ORCID

W Iff @ https://orcid.org/0000-0001-6465-8635

References

[1] Turunen J and Wyrowski F 1998 Diffractive Optics for Industrial and Commercial Applications (New York: Wiley)
[2] Lehr D, Helgert M, Sundermann M, Morhard C, Pacholski C, Spatz J P and Brunner R 2010 Opt. Express 18 23878–90
[3] Nevière M and Popov E 2003 Light Propagation in Periodic Media (Boca Raton, FL: CRC Press)
[4] Popov E 2012 Gratings: Theory and Numeric Applications (Marseille: Institut Fresnel)
[5] Shcherbakov A A and Tishchenko A V 2012 J. Quant. Spectrosc. Radiat. Transfer 113 2424–30
[6] Sheng Y, Feng D and Larochelle S 1997 J. Opt. Soc. Am. A 14 1562–8
[7] Layet B and Taghizadeh M R 1997 J. Opt. Soc. Am. A 14 1554–61
[8] Martin O J F and Piller N B 1998 Phys. Rev. E 58 3909–15
[9] Dechamps N, de Beaucoudrey N, Bourlier C and Toutsian S 2006 J. Opt. Soc. Am. A 23 359–69
[10] Collino F, Ghanemi S and Joly P 2000 Comput. Method. Appl. Mech. Eng. 184 171–211
[11] Schäidle A, Zschiedrich L, Burger S, Klose R and Schmidt F 2007 J. Comput. Phys. 226 477–93
[12] Magath T and Serebryannikov A E 2005 J. Opt. Soc. Am. A 22 2405–18
[13] van Beurden M C 2011 J. Opt. Soc. Am. A 28 2269–78
[14] van Beurden M C 2012 Prog. Electromagn. Res. B 36 133–49
[15] Shcherbakov A A and Tishchenko A V 2010 Quantum Electron. 40 538–44
[16] Shcherbakov A A and Tishchenko A V 2012 J. Quant. Spectrosc. Radiat. Transfer 113 158–71
[17] Porter B, Pardo F, Bouchon P, Halárd R and Pelouard J-L 2013 J. Opt. Soc. Am. A 30 573–81
[18] Tishchenko A V and Ifs W 2016 Procédé de calcul numérique de la diffraction d’une structure Patent WO2016135406A1
[19] Ifs W 2016 Rigorous fourier methods based on numerical integration for the calculation of diffractive optical systems PhD Thesis University of Erlangen-Nuremberg
[20] Petit R 1980 Electromagnetic Theory of Gratings (Berlin: Springer)
[21] Li L 1996 J. Opt. Soc. Am. A 13 1024–35
[22] Kim H, Lee I-M and Lee B 2007 J. Opt. Soc. Am. A 24 2513–27
[23] Bremmer H 1951 Commun. Pure Appl. Math. 4 105–15
[24] Pai D M and Awada K A 1991 J. Opt. Soc. Am. A 8 755–62
[25] Nevière M and Popov E 1992 J. Opt. Soc. Am. A 9 2095–6
[26] Nevière M and Montiel F 1994 Opt. Commun. 108 1–7
[27] Ko D Y K and Inson J C 1988 Phys. Rev. B 38 9945–51
[28] Ko D Y K and Sambles J R 1988 J. Opt. Soc. Am. A 5 1863–6
[29] Chateau N and Hugonin J-P 1994 J. Opt. Soc. Am. A 11 1321–31
[30] Maystre D 1994 Pure Appl. Opt. 3 975–93
[31] Cotter N P K, Preisit T W and Sambles J R 1995 J. Opt. Soc. Am. A 12 1097–103
[32] Bellman R and Kalaba R 1959 J. Math. Mech. 8 683–704
[33] Atkinson F V 1960 J. Math. Anal. Appl. 1 255–76
[34] Berk H L, Book D L and Pitsch D 1967 J. Math. Phys. 8 1611–9
[35] Kay I 1961 J. Math. Anal. Appl. 3 40–9
[36] Keller H B and Keller J B 1962 J. Soc. Ind. Appl. Math. 10 246–59
[37] Broer L J F 1963 Appl. Sci. Res. B 10 110–8
[38] Sluijter F W 1969 J. Math. Anal. Appl. 27 282–302
[39] Li L 1994 J. Opt. Soc. Am. A 11 2829–36
[40] Saad Y and Schultz M H 1986 SIAM J. Sci. Stat. Comput. 7 856–69
[41] Hanke-Bourgeois M 2009 Grundlagen der Numerischen Mathematik und des Wissenschaftlichen Rechnens 3rd edition (Wiesbaden: Vieweg + Teubner Verlag)
[42] Schwarz H A 1870 Vierteljahresschrift der Naturforschenden Gesellschaft in Zürich 15 272–86
[43] Ifs W, Lindlein N and Tishchenko A V 2014 J. Opt. 16 102004
[44] Shcherbakov A A and Tishchenko A V 2016 Generalized source method in curvilinear coordinates for 2D grating diffraction simulation J. Quant. Spectrosc. Radiat. Transfer 187 76–96
[45] Dahmen W and Reusken A 2006 Numerical Analysis of 2nd edition (Cambridge: Cambridge University Press) ch 5
[46] Horn R A and Johnson C R 2013 Matrix Analysis 2nd edition (Cambridge: Cambridge University Press) ch 5
[47] Knopp K 1990 Theory and Application of Infinite Series (New York: Dover)