

# A Linear-Complexity Layer-Coupling Algorithm for 1D- and 2D-Periodic Scattering in Multilayered Media

Loes F. van Rijswijk<sup>1, 2</sup>, Frank J. Buijnsters<sup>2</sup>, and Martijn C. van Beurden<sup>1, \*</sup>

**Abstract**—The formulation of a matrix-vector product with linear complexity for layer-coupling is discussed in the context of scattering by periodic dielectric scatterers embedded in a layered medium and formulated as a spectral-domain volume integral. It is shown how a traditional formulation in terms of reflection and transmission coefficients can be modified to arrive at an algorithm of linear complexity if being used as a matrix-vector product. The computational performance scheme is demonstrated for stacks in which scattering objects are distributed over hundreds of layers.

## 1. INTRODUCTION

The developments in the technological sector, and in particular the growth in data traffic, is reaching the limits of traditional NAND flash memory and drives the development of the next generation chip. A new type of flash memory was developed, called the V-NAND or 3D-NAND, which is a vertically stacked NAND. The 3D-NAND design improves memory capacity, speed, endurance, and efficiency [1]. The difference between these next generation chips and the traditional NAND flash memory chips lies in the number of layers of materials. Traditional NAND flash memory is on the order of ten layers, where 3D-NAND has hundreds of layers. In 2020, Micron and SK Hynix announced a 176-layer NAND design [2, 3]. The Imec technological roadmap anticipates 512 layer designs by 2022–2023 [4].

Because of this development, new challenges occur when these new chips are manufactured. Optical metrology has established itself as a powerful technology to control the IC fabrication process and further drive down tolerances, contributing to the decades-long trend of exponential growth in the computational power of devices (Moore’s law) [5]. Efficient electromagnetic solvers are needed to understand the sub-resolution scatterometric optical signal formation in device structures, which is essential in the development of hardware and algorithmic solutions for optical metrology, grating-profile reconstruction, and in the design of metrology targets optimized for optical control.

The electromagnetic modeling of 3D-NAND stacks, which like other memory devices are highly repetitive structures, requires a fast Maxwell solver that can handle a high level of complexity in the vertical direction (thick stack, many layers). There are a number of well-known techniques for computing electromagnetic scattering by periodic structures. These include the Finite-Difference Time Domain Method (FDTD) [6], the Rigorous Coupled Wave Analysis (RCWA) [7–9], the Surface Integral Method (SIM) [10, 11], and the Volume Integral Method (VIM) [12–16]. In this paper, we focus on the spectral formulation of the VIM, which offers an excellent speed–accuracy trade-off while being highly scalable, with a matrix-vector product of  $O(N \log N)$  in the number of Fourier modes used to represent the solution in the plane. The VIM works by separating a two- or three-dimensional system into a one-dimensional reference system (the layered background) and a contrast (the scatterers). The method relies on the fact that the homogeneous solutions and Green function of the reference system can be solved analytically and evaluated efficiently. We extend the applicability of the spectral VIM to stacks with a very large number of layers in the vertical direction by improving the scalability of a core

---

*Received 7 August 2022, Accepted 7 September 2022, Scheduled 6 October 2022*

\* Corresponding author: Martijn Constant van Beurden (M.C.v.Beurden@tue.nl).

<sup>1</sup> Eindhoven University of Technology, The Netherlands. <sup>2</sup> ASML Netherlands B.V., The Netherlands.

component of its algorithm. We demonstrate an improvement in computational complexity from  $O(L^2)$  to  $O(L)$ , where  $L$  is the number of layers. We expect the improvement to be relevant to the spectral VIM, as described in Section 2, as well as other formulations that rely on an efficient evaluation of the matrix-vector product corresponding to the Green function operator for layered media [17, 18].

This paper is organized as follows. In Section 2, we first introduce the general setup and introduce the expansion in the spectral domain. We then briefly discuss the structure of the Green function in a layered medium, as the key element in the VIM, and explain which part of the algorithm we propose to modify. Then, the spectral domain transmission-line equations and the interface conditions for layered media are derived, which we subsequently make explicit for the case of isotropic homogeneous layers. Section 3 explains the spectral integral representation and Green function. Two approaches for the evaluation of the integral representation are discussed, a traditional approach and a method that has linear computational complexity with respect to the number of layers. This section concludes with the spectral volume integral equation. Section 4 provides numerical examples comparing the traditional method and the linear-complexity method in terms of run-time performance. Furthermore, the improvement in case of the linear-complexity method is shown in case of a stack that can be compared to a realistic 3D-NAND stack. Finally, Section 5 states the conclusions.

## 2. SPECTRAL DOMAIN FORMULATION FOR LAYERED MEDIA

Consider a layered medium consisting of  $L$  dielectric layers, indicated by an index  $\ell = 1, \dots, L$ , enclosed between two half-spaces, indicated by  $\ell = 0$  for the top half-space and index  $\ell = L + 1$  for the bottom half-space. The stratification of the medium is perpendicular to the  $z$  direction, and the positions of the interfaces between the layers are indicated by  $Z_\ell$ . The medium per layer of the background is assumed to be homogeneous and isotropic, i.e., the relative permittivity is given by

$$\varepsilon_{b,r}(z) = \begin{cases} \varepsilon_{r,0}, & 0 = Z_0 < z, \\ \varepsilon_{r,i}, & Z_i < z < Z_{i-1}, i = 1, \dots, L, \\ \varepsilon_{r,L+1}, & z < Z_L. \end{cases} \quad (1)$$

We also define  $\varepsilon_b(z) = \varepsilon_0 \varepsilon_{b,r}(z)$ . Further, we assume nonmagnetic materials, i.e.,  $\mu = \mu_0$ . Parallel to the  $xy$  plane we assume a periodic arrangement of embedded dielectric scatterers with relative permittivity  $\varepsilon_r(\mathbf{r}_T, z)$ , where  $\mathbf{r}_T$  is the position vector in the  $xy$  plane, such that  $\varepsilon_r(\mathbf{r}_T + n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2, z) = \varepsilon_r(\mathbf{r}_T, z)$  with  $n_1$  and  $n_2$  integers and  $\mathbf{a}_1$  and  $\mathbf{a}_2$  the Bravais lattice vectors.

To formulate a volume integral equation for scattering of an incident plane wave by a periodic arrangement of dielectric scatterers embedded in the layered medium, we assume a  $\exp(j\omega t)$  time dependence, and we decompose the electromagnetic field in terms of the incident field and scattered field, where the incident electric and magnetic fields, denoted by  $\mathbf{E}^i$  and  $\mathbf{H}^i$ , satisfy Maxwell's equations in the absence of the dielectric scatterers. Consequently, the scattered fields, with superscript  $s$ , satisfy

$$\nabla \times \mathbf{H}^s = \mathbf{J} + j\omega \varepsilon_b(z) \mathbf{E}^s, \quad (2)$$

$$\nabla \times \mathbf{E}^s = -j\omega \mu_0 \mathbf{H}^s, \quad (3)$$

where  $\mathbf{J} = j\omega[\varepsilon(\mathbf{r}_T, z) - \varepsilon_0 \varepsilon_{b,r}(z)](\mathbf{E}^i + \mathbf{E}^s)$  is the contrast current density.

The incident plane wave is given by

$$\mathbf{E}^i = \mathbf{e}^i(z) \exp(-j\mathbf{k}_T^i \cdot \mathbf{r}_T), \quad (4)$$

with  $\mathbf{k}_T^i$  being the incident transverse wave vector, and  $\mathbf{e}^i(z)$  is such that it satisfies Maxwell's equations in for the layered-medium background. The scattered electric field is written in the spectral domain as

$$\mathbf{E}^s = \sum_{m_1=-\infty}^{\infty} \sum_{m_2=-\infty}^{\infty} \mathbf{e}^s(\mathbf{m}, z) \exp(-j\mathbf{k}_T^{\mathbf{m}} \cdot \mathbf{r}_T), \quad (5)$$

where  $\mathbf{m} = (m_1, m_2)$  and  $\mathbf{k}_T^{\mathbf{m}} = \mathbf{k}_T^i + m_1 \mathbf{b}_1 + m_2 \mathbf{b}_2$ , and  $\mathbf{b}_1$  and  $\mathbf{b}_2$  are the reciprocal lattice vectors. An analogous expansion is used for  $\mathbf{H}^s$  and  $\mathbf{J}$  in terms of the quantities  $\mathbf{h}^s(\mathbf{m}, z)$  and  $\mathbf{j}(\mathbf{m}, z)$ , respectively. The spectral expansion for periodic scattering is a sampled version of the formulation for a-periodic scattering, which can be found in, e.g., [17–19], and a similar procedure is used below to set the notation and explain how to arrive at the Green function and the corresponding integral equation for scattering by periodic dielectric objects, see also, e.g., [15].

## 2.1. General Strategy

In order to arrive at a scalable scheme, we make use of an iterative solution method to solve a volume-integral reformulation of the system (2)–(3), as will be discussed in more detail in Section 3. Thus, the performance of the scheme depends critically on the efficient implementation of two matrix-vector products, corresponding to (a) a spatial multiplication by the contrast function  $\varepsilon_r - \varepsilon_{b,r}$  and (b) the operator that involves the Green function  $\underline{\mathbf{G}}(\mathbf{m}, z, z')$  of the layered-medium background system.

The contrast operator ( $\varepsilon_r - \varepsilon_{b,r}$ ) is a spatial multiplication operator. Hence, when represented on the Fourier basis (5), its matrix-vector product has a block-Toeplitz structure and can be implemented with an  $O(N \log N)$  computational complexity. The scheme outlined in [16] is designed such that this operation can be carried out separately for each discretization point (sample) in  $z$ , though alternative approaches are possible.

The operator that involves the Green function couples the  $z$ -samples but does not mix the solutions for different mode indices  $\mathbf{m}$ . The evaluation of the matrix-vector product of the Green function operator, for a given mode index  $\mathbf{m}$  and contrast  $\mathbf{j}(\mathbf{m}, z)$ , consists of two steps:

- (i) evaluation of the solution  $\mathbf{e}^s$  inside each homogeneous background layer for the given contrast function  $\mathbf{j}(\mathbf{m}, z)$  (i.e., the particular part), up to the complex integration constants; and
- (ii) matching of the integration constants (the homogeneous parts) inside all layers to satisfy the continuity conditions at all layer interfaces. The corresponding set of conditions can be collected in a matrix equation, which takes the final form of (34)–(35) below.

Notice that step (i) corresponds to the convolution of the source term  $\mathbf{j}(\mathbf{m}, z)$  by the Green function for a homogeneous material, which depends only on the distance between source and observation points. This can be performed in  $O(N)$  or  $O(N \log N)$  time, where  $N$  is the number of sample points in  $z$ , provided that we make use of a suitable basis set [20, 21] to represent the functions inside the layer.

The focus of this paper is on the efficient implementation of step (ii). In the traditional approach, the integration constants are evaluated for each source layer separately. In other words, after the particular solution for one layer is evaluated, we extend this solution to all other layers, which are at that point assumed to contain no sources, using the interface conditions to fix the integration constants. In the end, the extended solutions for all source layers are added up according to the superposition principle. The total computation time is then  $O(L^2)$  in the number of layers  $L$ . Subsequently, we present an improved scheme that reduces this computational complexity to  $O(L)$ . In this approach, all particular solutions are computed first per layer, and their respective values at the layer interfaces are evaluated. The integration constants are then computed for all layers simultaneously in linear time by solving a banded-matrix system representing the interface conditions.

## 2.2. Transmission-Line Equations and Interface Conditions

Owing to the isotropic nature of the dielectric layers in the background, the scattered modal electric and magnetic fields,  $\mathbf{e}^s(\mathbf{m}, z)$  and  $\mathbf{h}^s(\mathbf{m}, z)$ , can be further decomposed according to the polarization states in parallel, perpendicular, and  $z$  components. This decomposition can be extended to the modal expansion of the contrast current density  $\mathbf{j}(\mathbf{m}, z)$  as

$$\mathbf{e}^s(\mathbf{m}, z) = e_{\parallel}(\mathbf{m}, z)\mathbf{u}_k + e_{\perp}(\mathbf{m}, z)(\mathbf{u}_z \times \mathbf{u}_k) + e_z(\mathbf{m}, z)\mathbf{u}_z, \quad (6)$$

$$\mathbf{h}^s(\mathbf{m}, z) = h_{\perp}(\mathbf{m}, z)\mathbf{u}_k + h_{\parallel}(\mathbf{m}, z)(\mathbf{u}_z \times \mathbf{u}_k) + h_z(\mathbf{m}, z)\mathbf{u}_z, \quad (7)$$

$$\mathbf{j}(\mathbf{m}, z) = j_{\parallel}(\mathbf{m}, z)\mathbf{u}_k + j_{\perp}(\mathbf{m}, z)(\mathbf{u}_z \times \mathbf{u}_k) + j_z(\mathbf{m}, z)\mathbf{u}_z, \quad (8)$$

where  $\mathbf{u}_z$  is the unit vector in the  $z$  direction and  $\mathbf{u}_k = \mathbf{k}_T^m / k_T^m$ , where  $k_T^m$  denotes the length of  $\mathbf{k}_T^m$ . This decomposition separates Maxwell's equations into two sets of coupled equations, i.e.,

$$\begin{cases} \frac{d}{dz}e_{\perp}(\mathbf{m}, z) = j\omega\mu_0 h_{\perp}(\mathbf{m}, z), \\ j\omega\mu_0 \frac{d}{dz}h_{\perp}(\mathbf{m}, z) = \gamma^2(\mathbf{m}, z)e_{\perp}(\mathbf{m}, z) + j\omega\mu_0 j_{\perp}(\mathbf{m}, z), \end{cases} \quad (9)$$

$$h_z(\mathbf{m}, z) = \frac{k_T^m e_{\perp}(\mathbf{m}, z)}{\omega\mu_0}, \quad (10)$$

$$\begin{cases} \frac{d}{dz}h_{\parallel}(\mathbf{m}, z) = -j\omega\varepsilon_b(z)e_{\parallel}(\mathbf{m}, z) - j_{\parallel}(\mathbf{m}, z), \\ j\omega\varepsilon_b(z)\frac{d}{dz}e_{\parallel}(\mathbf{m}, z) = -\gamma^2(\mathbf{m}, z)h_{\parallel}(\mathbf{m}, z) + jk_T^2j_z(\mathbf{m}, z) \end{cases}, \quad (11)$$

$$e_z(\mathbf{m}, z) = \frac{-k_T^{\mathbf{m}}h_{\parallel}(\mathbf{m}, z) + j[j_z(\mathbf{m}, z)]}{\omega\varepsilon_b(z)}, \quad (12)$$

where  $\gamma^2(\mathbf{m}, z) = (k_T^{\mathbf{m}})^2 - \omega^2\varepsilon_b(z)\mu_0$ , and its square root  $\gamma(\mathbf{m}, z)$  is taken according to the standard branch-cut definition, in line with the assumed time convention.

From now onwards, we will implicitly assume the dependence  $\mathbf{m}$  and will suppress  $\mathbf{m}$  in the notations of the various quantities to alleviate the notation. For each of the two sets of ordinary differential equations (ODEs) in  $z$  we obtain, per layer, two homogeneous solutions and a particular solution for each source type, i.e.,  $j_{\perp}$ ,  $j_{\parallel}$ , and  $j_z$ . These solutions per layer lead to the global solution (per mode  $\mathbf{m}$ ) by enforcing the interface conditions at each of the interfaces between the layers, i.e., by enforcing the continuity of  $e_{\perp}$ ,  $e_{\parallel}$ ,  $h_{\perp}$ , and  $h_{\parallel}$ . Per mode  $\mathbf{m}$ , we indicate the two homogeneous solutions in layer  $\ell$ , per polarization by

$$\mathbf{w}_{\perp,\ell}^+(z) = \begin{pmatrix} e_{\perp}^+(z) \\ h_{\perp}^+(z) \end{pmatrix} \quad \mathbf{w}_{\perp,\ell}^-(z) = \begin{pmatrix} e_{\perp}^-(z) \\ h_{\perp}^-(z) \end{pmatrix}, \quad (13)$$

$$\mathbf{w}_{\parallel,\ell}^+(z) = \begin{pmatrix} e_{\parallel}^+(z) \\ h_{\parallel}^+(z) \end{pmatrix} \quad \mathbf{w}_{\parallel,\ell}^-(z) = \begin{pmatrix} e_{\parallel}^-(z) \\ h_{\parallel}^-(z) \end{pmatrix}, \quad (14)$$

for  $z \in (Z_{\ell}, Z_{\ell-1})$ , where  $e_{\perp}^+$ ,  $h_{\perp}^+$ ,  $e_{\perp}^-$ ,  $h_{\perp}^-$  satisfy the set of ODEs in (9) in the absence of  $j_{\perp}$ , and  $e_{\parallel}^+$ ,  $h_{\parallel}^+$ ,  $e_{\parallel}^-$ ,  $h_{\parallel}^-$  satisfy the set of ODEs in (11) in the absence of  $j_{\parallel}$  and  $j_z$ . The superscripts denote the direction of propagation or decay of each of the solutions, i.e.,  $\mathbf{w}^+$  denotes an upward-traveling wave, and  $\mathbf{w}^-$  denotes a downward-traveling wave. We normalize the homogeneous solutions such that  $e_{\perp}^+(Z_{\ell}) = e_{\parallel}^+(Z_{\ell}) = 1$  and  $e_{\perp}^-(Z_{\ell-1}) = e_{\parallel}^-(Z_{\ell-1}) = 1$ .

If layer  $\ell$  also contains one or more source contributions, the particular solutions, restricted to layer  $\ell$ , are denoted by

$$\mathbf{p}_{\perp,\ell}[j_{\perp}](z) = \begin{pmatrix} e_{\perp}[j_{\perp}](z) \\ h_{\perp}[j_{\perp}](z) \end{pmatrix} \quad \mathbf{p}_{\parallel,\ell}[j_{\parallel}, j_z](z) = \begin{pmatrix} e_{\parallel}[j_{\parallel}, j_z](z) \\ h_{\parallel}[j_{\parallel}, j_z](z) \end{pmatrix}, \quad (15)$$

which are (linear) functionals of the source contributions, as indicated by the arguments between the square brackets, and  $z \in (Z_{\ell}, Z_{\ell-1})$ . In general, particular solutions can be obtained from the homogeneous solutions by the method of variation of parameters, see, e.g., [22], or by resorting to an equivalent transmission line and the relevant Green function pertaining to a suitable lumped generator [19, Section 2.4]. With these notations, we can write the solutions for the two polarizations in layer  $\ell$  as

$$A_{\perp,\ell}^- \mathbf{w}_{\perp,\ell}^-(z) + A_{\perp,\ell}^+ \mathbf{w}_{\perp,\ell}^+(z) + \mathbf{p}_{\perp,\ell}[j_{\perp}](z), \quad (16)$$

$$A_{\parallel,\ell}^- \mathbf{w}_{\parallel,\ell}^-(z) + A_{\parallel,\ell}^+ \mathbf{w}_{\parallel,\ell}^+(z) + \mathbf{p}_{\parallel,\ell}[j_{\parallel}, j_z](z). \quad (17)$$

By introducing the four  $2 \times 2$  boundary matrices at the top and bottom interfaces as

$$\underline{\underline{\mathbf{B}}}_{\perp,\ell}^{\uparrow} = \begin{pmatrix} \mathbf{w}_{\perp,\ell}^-(Z_{\ell-1}) & \mathbf{w}_{\perp,\ell}^+(Z_{\ell-1}) \end{pmatrix}, \quad \underline{\underline{\mathbf{B}}}_{\perp,\ell}^{\downarrow} = \begin{pmatrix} \mathbf{w}_{\perp,\ell}^-(Z_{\ell}) & \mathbf{w}_{\perp,\ell}^+(Z_{\ell}) \end{pmatrix}, \quad (18)$$

$$\underline{\underline{\mathbf{B}}}_{\parallel,\ell}^{\uparrow} = \begin{pmatrix} \mathbf{w}_{\parallel,\ell}^-(Z_{\ell-1}) & \mathbf{w}_{\parallel,\ell}^+(Z_{\ell-1}) \end{pmatrix}, \quad \underline{\underline{\mathbf{B}}}_{\parallel,\ell}^{\downarrow} = \begin{pmatrix} \mathbf{w}_{\parallel,\ell}^-(Z_{\ell}) & \mathbf{w}_{\parallel,\ell}^+(Z_{\ell}) \end{pmatrix}, \quad (19)$$

and the vectors with the amplitudes of the upward and downward-traveling homogeneous solutions

$$\mathbf{A}_{\perp,\ell} = \begin{pmatrix} A_{\perp,\ell}^- \\ A_{\perp,\ell}^+ \end{pmatrix}, \quad \mathbf{A}_{\parallel,\ell} = \begin{pmatrix} A_{\parallel,\ell}^- \\ A_{\parallel,\ell}^+ \end{pmatrix}, \quad (20)$$

the set of interface conditions for each polarization, at the interface  $z = Z_{\ell-1}$  between the layers  $\ell - 1$  and  $\ell$ , is given by

$$\underline{\underline{\mathbf{B}}}_{\perp,\ell-1}^{\downarrow} \mathbf{A}_{\perp,\ell-1} - \underline{\underline{\mathbf{B}}}_{\perp,\ell}^{\uparrow} \mathbf{A}_{\perp,\ell} = -\mathbf{p}_{\perp,\ell-1}[j_{\perp}](Z_{\ell-1}) + \mathbf{p}_{\perp,\ell}[j_{\perp}](Z_{\ell-1}), \quad (21)$$

$$\underline{\underline{\mathbf{B}}}_{\parallel,\ell-1}^{\downarrow} \mathbf{A}_{\parallel,\ell-1} - \underline{\underline{\mathbf{B}}}_{\parallel,\ell}^{\uparrow} \mathbf{A}_{\parallel,\ell} = -\mathbf{p}_{\parallel,\ell-1}[j_{\parallel}, j_z](Z_{\ell-1}) + \mathbf{p}_{\parallel,\ell}[j_{\parallel}, j_z](Z_{\ell-1}). \quad (22)$$

For the top and bottom interfaces, the conditions are slightly altered, since we are concerned here with the scattered field, and therefore, no waves from infinity approaching the top and bottom interfaces are allowed, i.e.,  $A_{\perp,0}^- = A_{\parallel,0}^- = 0$  and  $A_{\perp,L}^+ = A_{\parallel,L}^+ = 0$ . After solving these two sets of equations for all layer interfaces, the amplitudes of the homogeneous solutions in each layer are obtained, and in principle the scattered electric for mode  $\mathbf{m}$ , i.e.,  $\mathbf{e}^s(\mathbf{m}, z)$ , due to a source  $\mathbf{j}(\mathbf{m}, z)$ , is obtained. Finally, we note that the incident field  $\mathbf{e}^i$  satisfies the same system of Equations (21) and (22) with  $\mathbf{m} = \mathbf{0}$ , for its perpendicular and parallel components, respectively.

### 2.3. Transverse Field Components Specific for Isotropic Homogeneous Layers

The above formulation is valid for both a homogeneous and an inhomogeneous permittivity per layer, but the permittivity can only depend on  $z$ . We now concentrate on the case in which the permittivity is homogeneous in each layer. In layer  $\ell$  we then have the background permittivity  $\varepsilon_{\ell} = \varepsilon_0 \varepsilon_{r,\ell}$ , which is now constant, and we have for each mode the complex propagation coefficient  $\gamma_{\ell} = \sqrt{k_T^2 - \omega^2 \varepsilon_{\ell} \mu_0}$ .

The homogeneous solutions  $\mathbf{w}^+$  and  $\mathbf{w}^-$  for both polarizations are given by

$$\mathbf{w}_{\perp,\ell}^+(z) = \exp(-\gamma_{\ell}(z - Z_{\ell})) \begin{pmatrix} 1 \\ -Y_{\perp,\ell} \end{pmatrix} \quad \mathbf{w}_{\perp,\ell}^-(z) = \exp(-\gamma_{\ell}(Z_{\ell-1} - z)) \begin{pmatrix} 1 \\ Y_{\perp,\ell} \end{pmatrix}, \quad (23)$$

$$\mathbf{w}_{\parallel,\ell}^+(z) = \exp(-\gamma_{\ell}(z - Z_{\ell})) \begin{pmatrix} 1 \\ -Y_{\parallel,\ell} \end{pmatrix} \quad \mathbf{w}_{\parallel,\ell}^-(z) = \exp(-\gamma_{\ell}(Z_{\ell-1} - z)) \begin{pmatrix} 1 \\ Y_{\parallel,\ell} \end{pmatrix}, \quad (24)$$

with  $Y_{\perp,\ell} = \gamma_{\ell}/j\omega\mu_0$  and  $Y_{\parallel,\ell} = -j\omega\varepsilon_{\ell}/\gamma_{\ell}$ , where we have introduced the extra minus sign in  $Y_{\parallel}$  on purpose, to have two similar sets of equations for both polarizations. By adopting  $Y_{\ell}$  as the generic notation for  $Y_{\perp,\ell}$  or  $Y_{\parallel,\ell}$ , the general form of the equations in (21) and (22) for the interface at  $Z_{\ell-1}$  is now given by

$$\begin{pmatrix} X_{\ell-1} & 1 & -1 & -X_{\ell} \\ Y_{\ell-1}X_{\ell-1} & -Y_{\ell-1} & -Y_{\ell} & Y_{\ell}X_{\ell} \end{pmatrix} \begin{pmatrix} A_{\ell-1}^- \\ A_{\ell-1}^+ \\ A_{\ell}^- \\ A_{\ell}^+ \end{pmatrix} = -\mathbf{p}_{\ell-1}(Z_{\ell-1}) + \mathbf{p}_{\ell}(Z_{\ell-1}), \quad (25)$$

with  $X_{\ell} = \exp(-\gamma_{\ell}(Z_{\ell-1} - Z_{\ell}))$ , and we have dropped the notations that indicate the polarization and the source dependence in the vectors  $\mathbf{p}(Z_{\ell-1})$ . For the particular solution, we employ the outward traveling fundamental solution for a homogeneous medium for a source in layer  $q$ , i.e.,

$$g_q[j_p](z) = -\frac{1}{2\gamma_q} \int_{Z_q}^{Z_{q-1}} \exp(-\gamma_q|z - z'|) j_p(z') dz', \quad (26)$$

where  $p \in \{\perp, \parallel, z\}$  indicates the direction of the current density, and the observation point  $z$  is restricted to layer  $q$ . A set of particular solutions for sources in layer  $q$  can now be written as

$$\mathbf{p}_{\perp,\ell}[j_{\perp}](z) = \begin{pmatrix} e_{\perp}[j_{\perp}](z) \\ h_{\perp}[j_{\perp}](z) \end{pmatrix} = \begin{pmatrix} j\omega\mu_0 \\ \frac{d}{dz} \end{pmatrix} g_q[j_{\perp}](z) \quad (27)$$

and

$$\mathbf{p}_{\parallel,\ell}[j_{\parallel}, j_z](z) = \begin{pmatrix} e_{\parallel}[j_{\parallel}, j_z](z) \\ h_{\parallel}[j_{\parallel}, j_z](z) \end{pmatrix} = \begin{pmatrix} \frac{\gamma_q^2}{j\omega\varepsilon_b} & -\frac{1}{j\omega\varepsilon_b} \frac{d}{dz} \\ -\frac{d}{dz} & 1 \end{pmatrix} \begin{pmatrix} g_q[j_{\parallel}](z) \\ -jk_T^2 g_q[j_z](z) \end{pmatrix}. \quad (28)$$



### 3. SPECTRAL INTEGRAL REPRESENTATION AND GREEN FUNCTION

With the analysis from Section 2, we can derive an integral representation for each mode  $\mathbf{m}$  of the scattered electric field in terms of the contrast current density for the same mode. In the spectral domain, this yields, see, e.g., Section 2 of [15],

$$\mathbf{e}^s(\mathbf{m}, z) = \int_{z' \in \mathbf{R}} \underline{\underline{\mathbf{G}}}(\mathbf{m}, z, z') \mathbf{j}(\mathbf{m}, z') dz', \quad (36)$$

where the integration domain  $\mathbf{R}$  refers to the entire real axis or to the smaller interval in  $z$  along which the scattering objects are present in the background medium. Further,  $\underline{\underline{\mathbf{G}}}(\mathbf{m}, z, z')$  is a  $3 \times 3$  matrix function that represents the layered-medium Green function for an electric contrast source. For the transverse electric-field components the solution for the scattered field in layer  $\ell$  is obtained from (16), and the  $z$  component of the electric field is found from (12). Consequently, the scattered field in layer  $\ell$  due to a source in layer  $q$ , denoted by  $\mathbf{e}_{\ell,q}^s$ , with  $q$  possibly different from  $\ell$ , is given by

$$\begin{aligned} \mathbf{e}^s(\mathbf{m}, z) = & \left( \delta_{\ell,q} e_{\perp} [\mathbf{u}_k \cdot \mathbf{j}(\mathbf{m})](z) + A_{\perp,\ell,q}^+ e_{\perp}^+(z) + A_{\perp,\ell,q}^- e_{\perp}^-(z) \right) \mathbf{u}_k \\ & + \left( \delta_{\ell,q} e_{\parallel} [(\mathbf{u}_z \times \mathbf{u}_k) \cdot \mathbf{j}(\mathbf{m}), 0](z) + A_{\parallel,\ell,q}^+ e_{\parallel}^+(z) + A_{\parallel,\ell,q}^- e_{\parallel}^-(z) \right) (\mathbf{u}_z \times \mathbf{u}_k) \\ & + \frac{-k_T^{\mathbf{m}}}{\omega \varepsilon_{\ell}} \left( \delta_{\ell,q} h_{\parallel} [0, \mathbf{u}_z \cdot \mathbf{j}(\mathbf{m})](z) + A_{\perp,\ell,q}^+ h_{\parallel}^+(z) + A_{\perp,\ell,q}^- h_{\parallel}^-(z) \right) \mathbf{u}_z + \frac{j}{\omega \varepsilon_{\ell}} \delta_{\ell,q} (\mathbf{u}_z \cdot \mathbf{j}(\mathbf{m}, z)) \mathbf{u}_z, \quad (37) \end{aligned}$$

where  $\delta_{\ell,q}$  is the Kronecker delta,  $\varepsilon_{\ell}$  the background permittivity of layer  $\ell$ , and  $z$  lies in the interval that spans layer  $\ell$ . The coefficients  $A_{\perp,\ell,q}^{\pm}$  and  $A_{\parallel,\ell,q}^{\pm}$  are obtained as the solution from the linear system (34) for each of the two polarizations, where the right-hand-side vector  $\mathbf{p}$  contains only nonzeros for the interfaces  $Z_{q-1}$  and  $Z_q$  for the sources in layer  $q$ . The scattered field due to all the sources distributed over several layers is then obtained by summing the contributions over the source layers.

The columns of the Green function in layer  $\ell$  corresponding to the Cartesian components of  $\mathbf{j}(\mathbf{m}, z')$  are now obtained from the above expression if we choose  $\mathbf{j}(\mathbf{m}) = \mathbf{u}_p \delta(z - z')$ , with  $\mathbf{u}_p$  being a unit vector in the direction  $p$  for  $p \in \{x, y, z\}$ . Since the particular solutions in the linear system (34) now depend on  $z'$ , the coefficients  $A_{\perp,\ell,q}^{\pm}$  and  $A_{\parallel,\ell,q}^{\pm}$  now also become functions of  $z'$ . Fortunately, this does not mean that the linear system needs to be solved for every value of  $z'$ , since this function behavior can be factorized out, as we will discuss below.

#### 3.1. Traditional Approach: Construction of the Green Function

Above, we have briefly outlined how to formally obtain the Green function from the homogeneous and particular solutions in each layer and for each source type. The bottom line is that we have to solve the system of Equation (34) for every layer  $q$  in which a source is present and that the obtained coefficients  $A_{\perp,\ell,q}^{\pm}$  and  $A_{\parallel,\ell,q}^{\pm}$  depend on the source position  $z'$ . However, since the particular solutions for a source in layer  $q$  only need to be evaluated at the top and bottom interfaces at  $Z_{q-1}$  and  $Z_q$ , we can exploit (29) and (30) to express the right-hand-side vector  $\mathbf{p}$  in (34) as a constant vector times the functions  $g_q[\delta(Z_{q-1} - z')](Z_{q-1})$  and  $g_q[\delta(Z_q - z')](Z_q)$ . This means that we only need to generate the solution for two independent vectors  $\mathbf{p}_q^+ g_q[\delta(Z_{q-1} - z')](Z_{q-1})$  and  $\mathbf{p}_q^- g_q[\delta(Z_q - z')](Z_q)$  per layer and per polarization, where  $\mathbf{p}_q^{\pm}$  is a vector with all zeros, except for the entry in which the source in layer  $q$  appears at  $Z_{q-1}$  or  $Z_q$ , where the entry is obtained from (29) and (30). Hence we solve the two systems

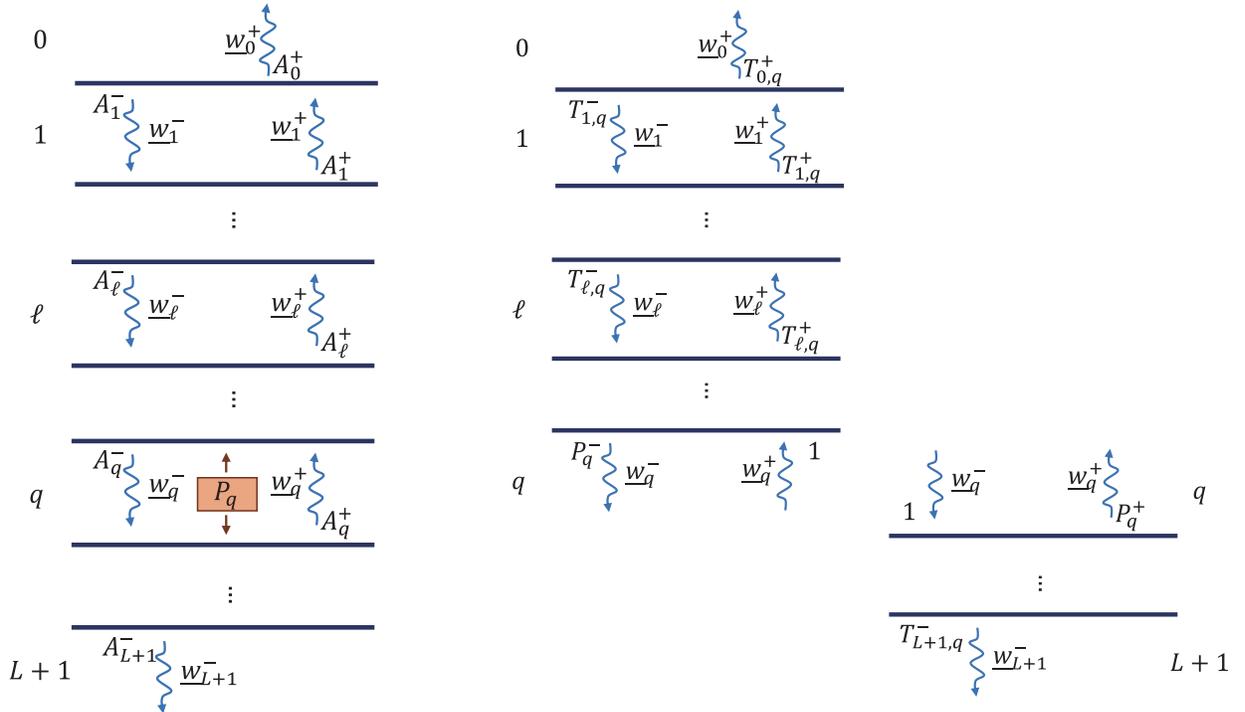
$$\underline{\underline{\mathbf{C}}}\mathbf{A}_{q,1}^{\pm} = \mathbf{p}_q^+, \quad \underline{\underline{\mathbf{C}}}\mathbf{A}_{q,2}^{\pm} = \mathbf{p}_q^-, \quad (38)$$

for both polarizations. For each polarization, the solution in layer  $\ell$  due to a point source in layer  $q \neq \ell$  at  $z = z'$  that satisfies all the boundary conditions is now obtained as

$$\begin{aligned} & \left\{ A_{\ell,q,1}^+ \exp(-\gamma_{\ell}(z - Z_{\ell})) + A_{\ell,q,1}^- \exp(-\gamma_{\ell}(Z_{\ell-1} - z)) \right\} g_q [\delta(Z_{q-1} - z')] (Z_{q-1}) \\ & + \left\{ A_{\ell,q,2}^+ \exp(-\gamma_{\ell}(z - Z_{\ell})) + A_{\ell,q,2}^- \exp(-\gamma_{\ell}(Z_{\ell-1} - z)) \right\} g_q [\delta(Z_q - z')] (Z_q), \quad (39) \end{aligned}$$

where the coefficients  $A_{\ell,q,i}^{\pm}$  with  $i \in \{1, 2\}$  are the entries of the solution vectors  $\mathbf{A}_{q,1}^{\pm}$  and  $\mathbf{A}_{q,2}^{\pm}$  pertaining to layer  $\ell$ . For the case  $\ell = q$ , the particular solution for layer  $q$  should be added extra. The advantage of this procedure is that the vectors  $\mathbf{A}_{q,1}^{\pm}$  and  $\mathbf{A}_{q,2}^{\pm}$  are now independent from  $z'$ , and therefore these vectors have to be computed only once. After that, the Green function can be written down explicitly. The structure of the Green function in terms of  $z$  and  $z'$  takes the form of a sum of four product kernels, i.e., multiplications of functions in  $z$  and  $z$  separately, which directly follows from (39).

If each layer contains a source, the above procedure samples all the columns of the inverse matrix  $\underline{\underline{C}}^{-1}$ . In case the number of layers containing a source is significantly smaller than the total number of layers in the background, it is usually more efficient to compute the solutions by exploiting the banded nature of the matrix  $\underline{\underline{C}}$ , for which the LU decomposition is also banded. Alternatively, and more traditionally, one could modify the matrix  $\underline{\underline{C}}$  in (34) by means of a rank-2 update that cancels the matrix elements  $S_{q-1,q}^{12}$  and  $S_{q,q+1}^{21}$ . The pertaining matrix then splits into two independent parts: one pertaining to the stack of layers above layer  $q$  and one for the stack of layers below layer  $q$ . These two systems can then be solved for a right-hand side vector with all zeros, except for a “1” at the top interface of layer  $q$  and a second right-hand side vector also containing all zeros but now a “1” at the bottom interface of layer  $q$ . In that case, the coefficients in the two obtained solution vectors for the separated linear systems can be interpreted as the reflection and transmission coefficients of the superstrate and substrate, as advocated in the construction of the spectral-domain Green function in, e.g., [17, 18]. These two solutions are then tied together by solving the remaining  $2 \times 2$  linear system that results from the initial rank-2 update of the matrix  $\underline{\underline{C}}$ . This procedure is graphically depicted in Fig. 1.



**Figure 1.** Coefficient definitions in the situation where  $q$  is the source layer,  $\ell$  the observation layer and there are no incoming fields. With the total stack on the left and on the right the stack is split, as explained in Section 3.1, into an upper part above the source and a lower part below the source.

Solving the system for the super- and substrate problems is traditionally solved efficiently and in a stable manner by means of the scattering matrix formalism and the Redheffer-star product [23] or via Riccati transformations [24], yet it is equivalent to computing the LU decomposition of both systems and employing forward-backward substitution.

We note that from a computational point of view, the product kernels obtained for the Green function yield an  $O(N_b + N_t)$  matrix-vector product between  $N_b$  basis functions in layer  $q$  and  $N_t$  testing functions in layer  $\ell$ . Since the functions forming the product kernels are independent of the precise combination of the layers, the homogeneous solutions in layer  $\ell$  in the decomposition can be reused for sources in other layers, and the particular solutions in layer  $q$  can be reused for other layers with observation points. However, the Green function itself still requires the evaluation for every combination  $\{\ell, q\}$  of observer and source layer, which for  $L$  layers represents a computational complexity of  $O(L^2)$  if each layer contains a source and a set of observation points.

### 3.2. An $O(L)$ Method for the Evaluation the Integral Representation

The idea behind the  $O(L)$  method for the evaluation of the integral representation is to avoid the computation of the reflection and transmission coefficients for each source layer separately. In this way, the Green function is not constructed explicitly, and only the total effect of all sources across all layers is computed for each observation layer.

Hence, instead of *a priori* solving for all  $\mathbf{A}^\pm$  for each combination of source and observation layers and tying these together as visualised in Fig. 1, the linear system in (34) is solved for all particular solutions simultaneously, i.e., for a single right-hand-side vector  $\mathbf{p}$  that contains the particular solutions in all layers. This means that matrix  $\underline{\underline{C}}$  needs to be solved only once, as long as the particular solutions remain unchanged. Owing to the banded-matrix structure of matrix  $\underline{\underline{C}}$ , the system in (34) can be solved using an optimized LU decomposition as described in Appendix A. This optimized decomposition is a specialized version of the LU decomposition proposed in [25], obtained by exploiting the additional features of the matrix  $\underline{\underline{C}}$ , i.e., the diagonal that contains only ones and the repeated presence of zeros on the outer two diagonals. The resulting LU factors can be obtained in  $O(L)$  steps, and the forward and backward substitution are also performed in  $O(L)$  steps. Consequently, this system can be solved up to machine precision by this direct matrix method in linear time. For repeatedly solving the linear system, as briefly discussed in the context of solving a volume integral equation in Section 3.3, the LU factors have to be computed only once, and only the forward and backward substitutions have to be performed repeatedly.

### 3.3. Spectral Volume Integral Equation

The integral representation in (36) can be employed as part of a volume integral equation between the modal total electric fields,  $\mathbf{e}(\mathbf{m}, z) = \mathbf{e}^i(\mathbf{m}, z) + \mathbf{e}^s(\mathbf{m}, z)$  and the constitutive relations between the electric field and the contrast current density. Due to the spectral nature, in which the Gibbs phenomenon at material interfaces is strong, we employ a formulation based on normal-vector fields [16, Sec. 2.2], i.e., we work with an intermediate modal field  $\mathbf{f}(\mathbf{m}, z)$  that is composed of the modal field of the electric field and the electric flux density. The resulting formulation is then given by the Lippmann-Schwinger equation

$$\begin{aligned} \mathbf{e}^i(\mathbf{m}, z) &= \mathbf{e}(\mathbf{m}, z) - \int_{z' \in \mathbf{R}} \underline{\underline{G}}(\mathbf{m}, z, z') \mathbf{j}(\mathbf{m}, z') dz', \\ \mathbf{e}(\mathbf{m}, z) &= (c_\varepsilon * \mathbf{f})(\mathbf{m}, z), \\ \mathbf{j}(\mathbf{m}, z) &= [(\varepsilon c_\varepsilon - \varepsilon_b(z) \cdot c_\varepsilon) * \mathbf{f}](\mathbf{m}, z), \end{aligned} \quad (40)$$

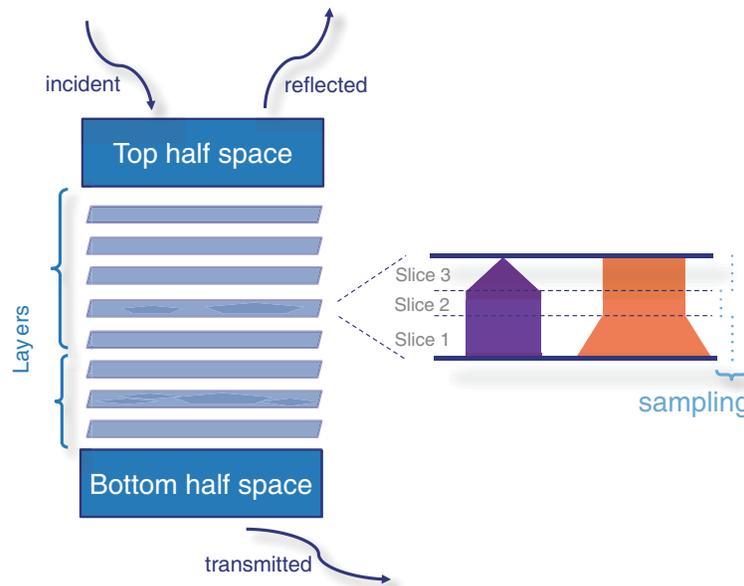
where the constitutive relations are point-wise multiplications in  $z$  and discrete convolutions in the modal index  $\mathbf{m}$ . The convolution operators  $c_\varepsilon *$  and  $(\varepsilon c_\varepsilon - \varepsilon_b(z) \cdot c_\varepsilon) *$  are the spectral counterparts of (12) and (14) in [16].

Equation (40) is a volume integral equation for the intermediate field  $\mathbf{f}(\mathbf{m}, z)$ . Owing to the discrete convolution structure of the constitutive relations and the nature of the Green function per layer, an efficient matrix-vector product can be obtained with computational complexity of  $O(MN \log(M))$  where  $N$  is the number of sample points in the  $z$  direction, and  $M$  is the total number of Fourier modes in the  $xy$  plane. Nevertheless, if we look at the complexity of the coupling between the  $L$  layers, a penalty in the form of  $O(L^2)$  is obtained if the traditional integral representation via the Green function is

used, as explained in Section 3.1. However, if the integral representation as explained in Section 3.2 is employed, the complexity reduces to  $O(L)$ , and the entire matrix-vector product is of linear complexity with respect to the  $z$  direction.

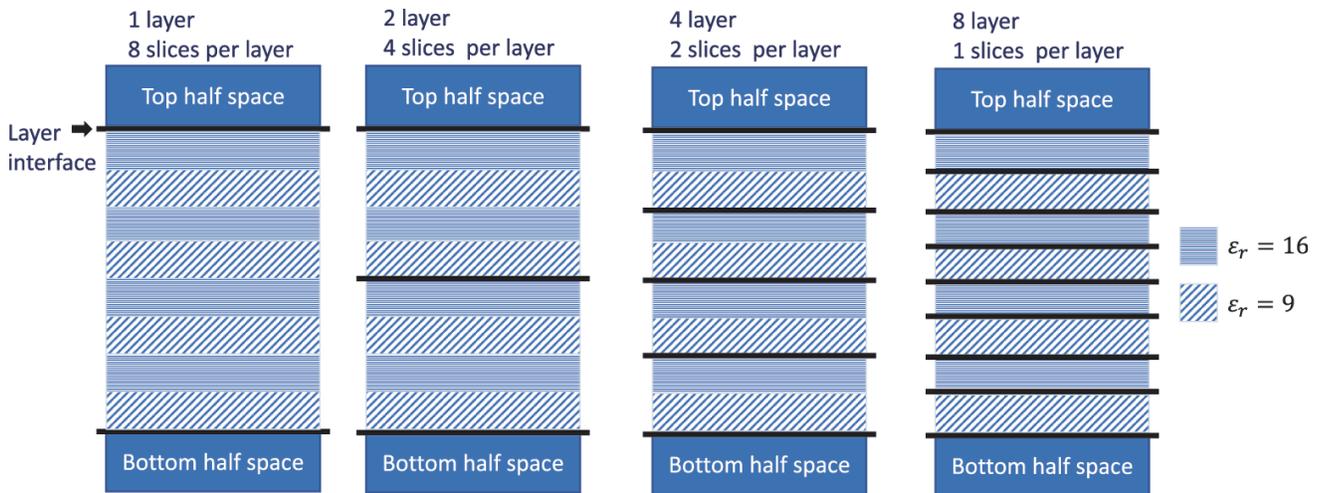
#### 4. NUMERICAL RESULTS

First we start with some terminology regarding the test stacks that we use in the section. A stack is a structure divided into layers. A layer contains a homogeneous background medium and is placed somewhere in between the two half spaces  $z > 0$  and  $z < Z_L$ , i.e., the top and bottom half spaces, respectively. A layer can contain one or more slices, each of which has a permittivity different from that of the background layer, i.e., it can contain dielectric objects with a different permittivity. A discretization for the contrast current density and for the total electric field is applied to every slice with a difference in permittivity compared to the permittivity of the background layer, i.e., the slice has a contrast. See Fig. 2 for a visual interpretation. In addition, we have an incident field with a wavelength in vacuum in the visible-light regime.

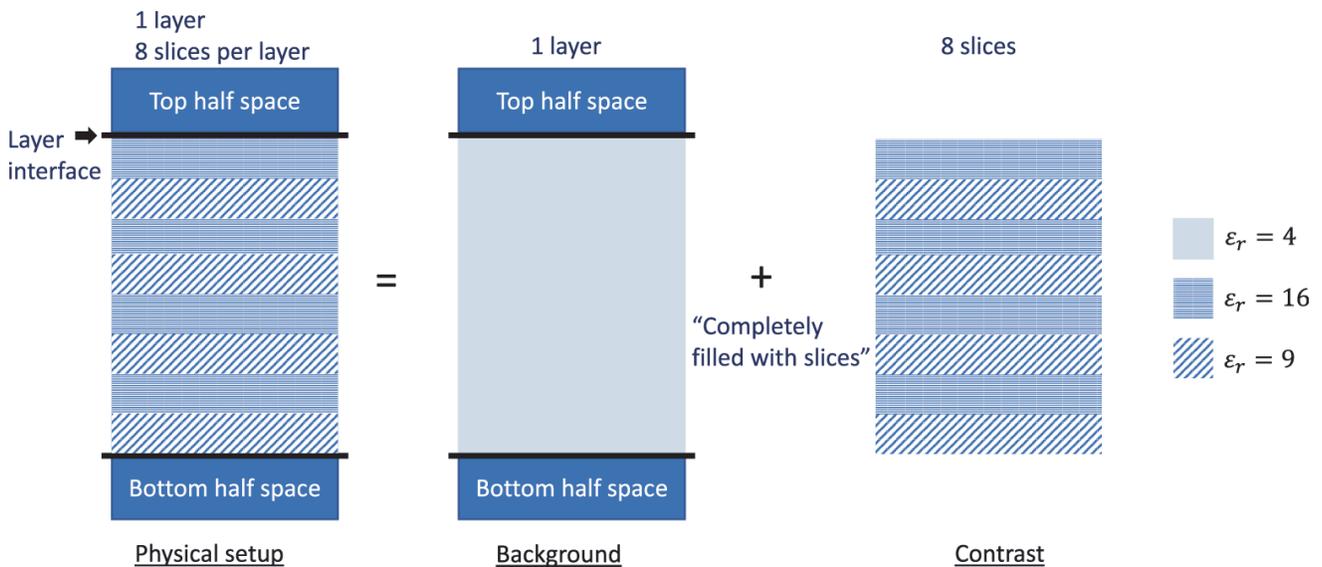


**Figure 2.** Interpretation stack terminology.

We show the difference in computation time to solve a volume integral equation using the traditional approach, i.e., by constructing and employing the Green function, as described in Section 3.1, and using the  $O(L)$  method as described in Section 3.2. To show the difference in computation time between the two methods, we define an artificial test stack referred to as Stack 1. Stack 1 is a stack containing 2048 slices while the number of layers and number of slices per layer are varied. The permittivity for each consecutive pair of slices is the same. Note that periodicity in the transverse direction is irrelevant since every slice contains a homogeneous material. For a visual interpretation of this partitioning see Fig. 3, where only 8 slices are partitioned into layers, for visual convenience. Every stack in this figure represents the same physical problem, i.e., only the division in layers and the number of slices in each layer are different. The background permittivity is the same for each layer but not shown separately in this figure. Fig. 4 shows a visual interpretation of Stack 1, and for convenience the same permittivities are used but now with fewer slices. The stack is visualised above “physical setup” and splits into the background layer and the contrast within each layer. This difference in partitioning results in differences in computation time because the coupling between layers is computed via the integral representation that contains interface conditions across each layer interface, whereas for slices within one layer the coupling takes place by means of the particular solution within one layer, see also [15]. Having the



**Figure 3.** Example stack varying from 1 layer containing 8 slices to 8 layers each containing one slice as an interpretation of the test case of Stack 1. All stacks represent the same physical problem.

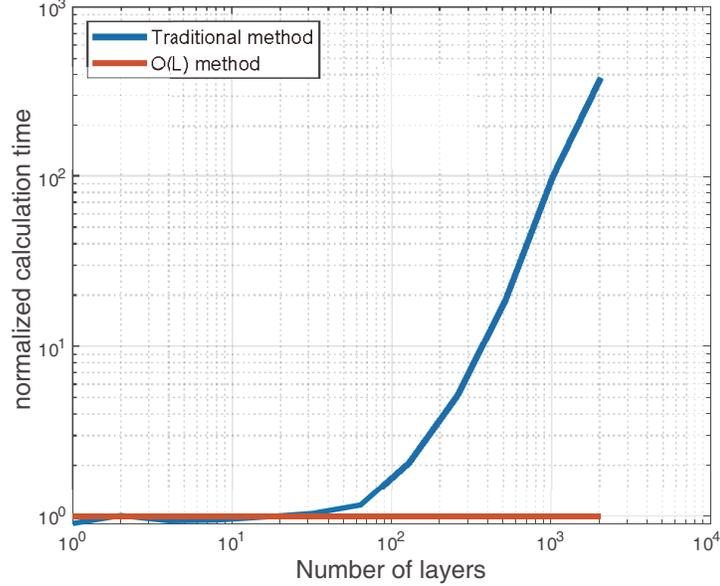


**Figure 4.** Interpretation of an example stack with profile characteristics as Stack 1. The layer is completely filled with slices. Varying the layers and slices per layer, in combination with the fixed background permittivity results in a test case for with all variations of the stack in terms of layers and slices for Stack 1 amount to the same physical problem.

same homogeneous background for all slices does not mean it can be skipped. The coupling between the layers is achieved by means of reflection and transmission coefficients of the entire stack, which are nonzero. In the case of Stack 1 with different partitioning of slices in layers, the computational difference between the traditional method and  $O(L)$  method is shown in Fig. 5. For completeness, we first mention that the details for the profile characteristic of Stack 1 are given in Table 1. Furthermore, for the simulations we use an incident plane wave at normal incidence, and since each slice contains only a homogeneous medium, the polarization is irrelevant. Fig. 5 shows the mean computation time over five identical simulations, to remove bias due to the state of the computer, for the test case of Stack 1 using the traditional approach to solve the volume integral equation, normalized to the computation time when the  $O(L)$  method is used. We observe that the quadratic behavior, i.e.,  $O(L^2)$  of the traditional

**Table 1.** Profile characteristics Stack 1.

Wavelength	1	permittivity odd slices	9
Thickness per slice	0.01	permittivity even slices	16
Background permittivity	4	permittivity top and bottom half spaces	1



**Figure 5.** Mean computation time over five simulations for the traditional approach to solve the volume integral equation for the  $O(L)$  method, normalized with respect to the  $O(L)$  method, for the situation of Stack 1. Stack 1 is a stack with real permittivities varying from 2048 slices in a single layer to one slice per layer with a total of 2048 layers.

approach with respect to the  $O(L)$  method, becomes visible around 100 layers. The  $O(L)$  method does not dominate the overall computation time in this example, which is why the quadratic behavior is observed for the traditional method.

We demonstrate the impact on computation time by solving the spectral volume integral equation in Section 3.3 with the  $O(L)$  method for more realistic cases. To this end, we define two test stacks in Table 2 and refer to them as Stack 2 and Stack 3. A visualization of the top twelve layers of Stack 2 and Stack 3 is shown in Fig. 6. Stack 2 and Stack 3 are designed in a similar way. The main difference

**Table 2.** Profile characteristics Stack 2 and Stack 3. Bilayer block is yellow-blue.

Characteristics	Stack 2	Stack 3
Nr. layers	219	75
Nr. cylindrical holes	216	72
Thickness total (nm)	7058	3016
Thickness bilayers (nm)	22–34	22–34
Permittivity bilayers	3.96–2.22	3.96–2.22
Wavelength (nm)	700	700
Periodicity in $x$ - $y$ (nm)	152.7–205.0	152.7–205.0
Diameter cylindrical holes (nm)	100	100



**Figure 6.** Visualisation of Stack 2 and 3, top 12 layers.

**Table 3.** Improvement factors in computation time for Stack 2 and Stack 3 when using the  $O(L)$  method to reconstruct the integral representation with respect to the traditional approach.

	Stack 2	Stack 3
Improvement factor total computation time	8	3
Improvement factor integral representation	27	9
Percentile change integral representation/total time	86% $\rightarrow$ 24%	64% $\rightarrow$ 20%

is that Stack 2 contains three times as many layers as Stack 3. We expect to see approximately the same factor in terms of computation time, which would reflect the  $O(L)$  behavior. Table 3 shows the results for Stack 2 and Stack 3 in terms of improvement factor in computation time for the  $O(L)$  method versus that of the traditional approach. For the simulations we used both parallel and perpendicular polarizations, and the coordinates of the angle of incidence of the plane wave are  $(\theta, \phi) = (35.7, -121.0)$ . Observe that the improvements in computation time related to Stack 2 are approximately a factor of 3 larger than Stack 3, as expected. When we focus on Stack 2, an improvement factor of 27 is obtained for the integral representation calculation, i.e., owing to the improved computational complexity of the  $O(L)$  method described in Section 3.2 compared to the traditional method described in Section 3.1. Furthermore, 86% of the total time was spent on the computation of the integral representation in the case of the traditional approach, whereas this is reduced to 24% for the  $O(L)$  method. In total, this results in an improvement factor of 8 in the case of Stack 2. Additionally, a significant improvement is made in terms of memory use in the case of a stack with many layers. The traditional method needs to have a full initialization of the matrix with reflection and transmission coefficients for the evaluation of every combination  $\{\ell, q\}$  of observer and source layer, which for  $L$  layers represents a full  $O(L \times L)$  matrix. The  $O(L)$  only needs to initialize the five diagonals related to the LU decomposition of matrix  $\underline{\underline{C}}$ . This results in an initialization of approximately  $5 \times 2 \times O(L)$ . In the case of Stack 2, this results in 12 GB gain in terms of memory use. So, apart from the improvements in computation time, an improvement factor of 7 is obtained in terms of memory for Stack 2. The above results are for one specific angle of incidence. We recall that both the homogeneous-medium part of the Green function in (26) and the matrix  $\underline{\underline{C}}$  in (35) change when the angle of incidence is changed, since they both depend on  $\mathbf{k}_T^i$  via  $\gamma_\ell$ . To show that the results are similar for other angles of incidence, a final set of results is given in Table 4, which shows the results for Stack 2 in the case of 200 randomly picked angles of incidence, for a short and long wavelength in the visible-light regime, i.e., 425 nm and 700 nm, for a fixed number of Floquet modes and number of basis functions in  $z$  for all angles. The table shows the total time, the specific time spent on the integral representation calculation, and the required number

**Table 4.** Computation time one specific angle of incidence (AOI) versus multiple AOI for two different wavelengths in case of Stack 2.

700 nm wavelength	Total time (s)	Time integral representation (s)	Nr. MVP
one specific AOI	223	54	66
average over 200 AOI	172	58	68
425 nm wavelength	Total time (s)	Time integral representation (s)	Nr. MVP
one specific AOI	208	47	63
average over 200 AOI	163	51	66

of matrix-vector products (MVP) to solve the integral equation. We observe that the timings related to a specific angle of incidence and the average time for multiple angles of incidence are approximately the same and the difference are due to the initialization time. In the case of multiple angles of incidence, the initialization pertaining to the geometry of the scatterers is required only once.

## 5. CONCLUSIONS

Starting from a spectral formulation of the volume integral method we extended to the case of a large number of layers by improving the scalability of the layer-coupling algorithm. We demonstrated an improvement in computational complexity from  $O(L^2)$  to  $O(L)$ , where  $L$  is the number of layers containing source and observation points, whereas we demonstrated that in the traditional approach the computation time of the volume integral equation is an  $O(L^2)$  algorithm. The traditional approach is efficient in the situation that  $L$  is small, since no calculations need to be repeated. All reflection and transmission coefficients are calculated once. The proposed improved algorithm is a linear method in terms of the number of layers containing source and observation points, and is very efficient in the situation where  $L$  is large, since coupling coefficients are solved for on the fly.

The next generation chips will contain many layers (large  $L$ ), meaning that a scattering analysis for such chips will benefit a lot from this algorithm. This statement is confirmed by the results. In terms of a realistic stack that can be compared to the current 3D-NAND, this resulted in an improvement factor of 27 in terms of the integral representation. As a consequence, an improvement factor of 8 in terms of overall computation time was achieved. Additionally, a significant improvement was made in terms of memory use, i.e., an improvement factor of 6 was obtained.

## APPENDIX A. SOLVING THE LINEAR SYSTEM OF THE O(L) METHOD

To solve the linear system given in (34) we use an optimized LU decomposition based on [25]. Using the special matrix form, i.e., the diagonal of all ones and the repeated presences of zeros on the upper and lower diagonals, the algorithm related to penta-diagonal linear systems is made somewhat more efficient. Below, the algorithm is given for the linear system  $\underline{\underline{C}}\mathbf{x} = \mathbf{p}$  given in (34), where for convenience we use the notation of  $\mathbf{x}$  instead of  $\mathbf{A}^\pm$ .

Decomposing matrix  $\underline{\underline{C}}$  into a lower triangular matrix  $\underline{\underline{L}}$  and an upper triangular matrix  $\underline{\underline{U}}$  results in

$$\overbrace{\begin{pmatrix} 1 & 0 & d_1 & & & \\ 0 & 1 & c_2 & & & \\ & b_3 & 1 & 0 & d_3 & \\ & a_4 & 0 & 1 & c_4 & \\ & & & b_5 & 1 & 0 \\ & & & a_6 & 0 & 1 \end{pmatrix}}^{\underline{\underline{C}}} \quad (\text{A1})$$

$$= \underbrace{\begin{pmatrix} 1 & & & & & & \\ g_2 & 1 & & & & & \\ 0 & b_3 & 1 & & & & \\ & a_4 & g_4 & 1 & & & \\ & & 0 & b_5 & 1 & & \\ & & & a_6 & g_6 & 1 & \end{pmatrix}}_{\underline{\underline{\mathbf{L}}}} \underbrace{\begin{pmatrix} e_1 & 0 & d_1 & & & & \\ & 1 & f_2 & & & & \\ & & e_3 & 0 & d_3 & & \\ & & & 1 & f_4 & & \\ & & & & e_5 & 0 & \\ & & & & & & 1 \end{pmatrix}}_{\underline{\underline{\mathbf{U}}}}, \tag{A2}$$

where

$$e_i = 1 - g_i f_{i-1}, \quad \text{for } i = 1, 3, 5, \dots, n \tag{A3a}$$

$$f_i = c_i - g_i d_{i-1}, \quad \text{for } i = 2, 4, \dots, n - 2 \tag{A3b}$$

$$g_i = -\frac{a_i f_{i-2}}{e_{i-1}}, \quad \text{for } i = 2, 4, \dots, n \tag{A3c}$$

with starting conditions  $f_0 = g_1 = 0$ . Now, the system  $\underline{\underline{\mathbf{L}}}\zeta = \mathbf{p}$  can be solved via forward substitution using the following equations

$$\zeta_i = \begin{cases} p_i, & \text{for } i = 1 \\ p_i - L_{1,i}\zeta_{i-1}, & \text{for } i = 2 \\ p_i - L_{1,i}\zeta_{i-1} - L_{2,i}\zeta_{i-2}, & \text{for } i = 3, 4, \dots, n \end{cases} \tag{A4}$$

and finally we solve for  $\underline{\underline{\mathbf{U}}}\mathbf{x} = \zeta$  via backward substitution with the equations

$$x_i = \begin{cases} \frac{\zeta_i - U_{1,i}x_{i+1} - U_{2,i}x_{i+2}}{U_{0,i}}, & \text{for } i = 1, \dots, n - 2 \\ \frac{\zeta_i - U_{1,i}x_{i+1}}{U_{0,i}}, & \text{for } i = n - 1 \\ \frac{\zeta_i}{U_{0,i}}, & \text{for } i = n. \end{cases} \tag{A5}$$

In the calculation of  $x_i$ , we only divide numerator by  $U_{0,i}$  when  $i$  is odd because  $U_{0,i} = 1$  when  $i$  is even. The first subscript of  $\underline{\underline{\mathbf{L}}}$  is related to its lower-diagonals (index 1 and 2), and the first subscript of  $\underline{\underline{\mathbf{U}}}$  is related to its diagonal (index 0) or lower-diagonals (index 1 and 2). The second subscript of  $\underline{\underline{\mathbf{L}}}$  and  $\underline{\underline{\mathbf{U}}}$  refer to the location of the element in the diagonal vector that the first subscript refers to.

## REFERENCES

1. Samsung Electronics Co., Ltd., “Samsung V-NAND technology”, Sept. 2014.
2. Micron, “176-Layer NAND,” <https://www.micron.com/products/nand-flash/176-layer-nand>, 2020, [Online, Accessed June 2021].
3. SK hynix newsroom, “SK hynix Unveils the Industry’s Most Multilayered 176-Layer 4D NAND Flash,” <https://news.skhynix.com/sk-hynix-unveils-the-industrys-highest-layer-176-layer-4d-nand-flash/>, 2020, [Online, Accessed June 2021].
4. Lapedus, M., “3D NAND flash Wars Begin,” *Semiconductor Engineering*, Aug. 2018. [Online, Accessed June 2021].
5. Maas, J., M. Ebert, K. Bhattacharyya, H. Cramer, A. Becht, S. Keij, R. Plug, A. Fuchs, M. Kubis, T. Hoogenboom, and V. Vaenkatesan. “YieldStar: A new metrology platform for advanced lithography control,” *Proc. SPIE. 7985, 27th European Mask and Lithography Conference*, 2011.
6. Yee, K. S., “Numerical solution of initial boundary value problems involving Maxwell’s equations in isotropic media,” *IEEE Transactions on Antennas and Propagation*, Vol. 14, No. 3, 302–307, 1966.
7. Moharam, M. G. and T. K. Gaylord, “Diffraction analysis of dielectric surface-relief gratings,” *J. Opt. Soc. Am.*, Vol. 72, No. 10, 1385–1392, 1982.

8. Li, L., "Use of Fourier series in the analysis of discontinuous periodic structures," *J. Opt. Soc. Am. A.*, Vol. 13, No. 9, 1870–1876, 1996.
9. Li, L., "New formulation of the Fourier modal method for crossed surface-relief gratings," *J. Opt. Soc. Am. A.*, Vol. 14, No. 10, 2758–2767, 1997.
10. Valerio, G., P. Baccarelli, S. Paulotto, F. Frezza, and A. Galli, "Regularization of mixed-potential layered-media Green's functions for efficient interpolation procedures in planar periodic structures," *IEEE Transactions on Antennas and Propagation*, Vol. 57, 122–134, 2009.
11. Jorna, P., V. Lancellotti, and M. C. van Beurden, "Computational aspects of 2D-quasi-periodic-Green-function computations for scattering by dielectric objects via surface integral Equations," *Progress In Electromagnetics Research PIER B*, Vol. 63, 49–66, 2015.
12. Chang, Y. C., G. Li, H. Chu, and J. Opsal, "Efficient finite-element Green's function approach for critical-dimension metrology of three-dimensional gratings on multilayer films," *J. Opt. Soc. Am. A*, Vol. 23, No. 3, 638–645, 2006.
13. Magath, T., "Coupled integral equations for diffraction by profiled, anisotropic, periodic structures," *IEEE Transactions on Antennas and Propagation*, Vol. 54, No. 2, 681–686, 2006.
14. Shi, Y. and C. H. Chan, "Multilevel Green's function interpolation method for analysis of 3-D frequency selective structures using volume/surface integral equation," *J. Opt. Soc. Am. A*, Vol. 27, 308–318, 2010.
15. Van Beurden, M. C., "Fast convergence with spectral volume integral equation for crossed block-shaped gratings with improved material interface conditions," *J. Opt. Soc. Am. A*, Vol. 28, No. 11, 2269–2278, 2011.
16. Van Beurden, M. C., "A spectral volume integral equation method for arbitrary bi-periodic gratings with explicit Fourier factorization," *Progress In Electromagnetic Research B*, Vol. 36, 133–149, 2012.
17. Michalski, K. A. and D. Zheng, "Electromagnetic scattering and radiation by surfaces of arbitrary shape in layered media, Part 1: Theory," *IEEE Transactions on Antennas and Propagation*, Vol. 38, No. 3, 335–344, 1990.
18. Michalski, K. A. and J. R. Mosig, "Multilayered media Green's functions in integral equation formulations," *IEEE Transactions on Antennas and Propagation*, Vol. 45, No. 3, 508–519, 1997.
19. Felsen, L. B. and N. Marcuvitz, *Radiation and Scattering of Waves*, Wiley-IEEE Press, 1994.
20. Gohberg, I. and I. Koltracht, "Numerical solution of integral equations, fast algorithms and Krein-Sobolev equation," *Numerical Mathematics*, Vol. 47, No. 2, 237–288, 1985.
21. Van Beurden, M. C., T. Zacharopoulou, A. Roc'h, and M. G. M. M. van Kraaij, "A pseudo-spectral longitudinal expansion in a spectral domain integral equation for scattering by periodic dielectric structures," *Proceedings of the 2015 International Conference on Electromagnetics in Advanced Applications (ICEAA)*, 1419–1422, Torino, Italy, September 7–11, 2015.
22. Kreyszig, E., *Advanced Engineering Mathematics*, John Wiley & Sons, 1988.
23. Ko, D. Y. K. and J.R. Sambles, "Scattering matrix method for propagation of radiation in stratified media: Attenuated total reflection studies of liquid crystals," *J. Opt. Soc. Am. A*, Vol. 5, No. 11, 1863–1866, 1988.
24. Van Kraaij, M. G. M. M., "Forward diffraction modelling: Analysis and application to grating reconstruction," Ph.D. thesis, Ch. 5, Eindhoven University of Technology, 2011.
25. Karawia, A., "A computational algorithm for solving periodic penta-diagonal linear systems," *Appl. Math. Comput.*, Vol. 174, No. 1, 613–618, 2006.