Accelerating the discrete dipole approximation by initializing with a scalar solution and using a circulant preconditioning

Patrick C Chaumet, Guillaume Maire, Anne Sentenac

Accelerating the discrete dipole approximation by initializing with a scalar solution and using a circulant preconditioning

Patrick C. Chaumet

Aix Marseille Univ, CNRS, Centrale Marseille, Institut Fresnel, Marseille, France

Guillaume Maire

Aix Marseille Univ, CNRS, Centrale Marseille, Institut Fresnel, Marseille, France

Anne Sentenac

Aix Marseille Univ, CNRS, Centrale Marseille, Institut Fresnel, Marseille, France

Abstract

The discrete dipole approximation (DDA) is a method of choice for simulating the electromagnetic scattering by objects of arbitrary shape and permittivity. To recover the field inside the object, it requires the iterative solving of a dense linear system which can be time consuming. To ease this task, we propose to start the inversion with the solution of the recently introduced scalar approximation [Chaumet *et al.* J. Opt. Soc . Am. A, **39**, 1462 (2022)]. This initial guess allows a reduction of the time required for the solving of the linear system up to 50%. In addition, we study the interest of preconditioning the system to accelerate convergence. We show that the gain can be up to a factor of 5, especially for homogeneous objects on a plane substrate.

Keywords: Scattering; Discrete dipole approximation; Iterative method, preconditioner;

1. Introduction

Simulating the interaction of an electromagnetic field with a three dimensional object of arbitrary shape and relative permittivity is mandatory in many

areas of photonics and surface science. There exist various techniques for solving Maxwell's equations in a complex environment in the time-harmonic regime [1, 2], including the finite element method (FEM), the multiple multipole method (MMP), the fast-multipole method (FMM), the method of moments (MoM) and the discrete dipole approximation (DDA).

In this article, we focus on the DDA, which is particularly adapted to the study of the light scattered by clusters of possibly anisotropic particles of arbitrary shapes [3, 4, 5, 6]. Reviews on the potentiality of DDA can be found in [7, 8], and many DDA codes are freely available on the net [9, 10, 11].

The DDA can be described as a two step process. The first step, which is the main bottleneck of the approach, consists in estimating the field existing inside the object by solving a self-consistent equation discretized into a dense linear system. The second step consists in computing the scattered field at the required observation points. The second step has been solved recently with a Fourier based method which permits a significant reduction of the computation time [12]. For the first step, combining the use of Fourier transforms (FFT) for the matrix vector products (MVP) [13] and iterative solvers [14, 15, 8, 16], has permitted to accelerate this calculation. Yet, for large objects, the computation time of the MVP becomes important and any means for reducing the number of iterations of the solvers is welcome. To this aim, different implementations of the DDA were investigated [17, 18]. Thus, using the filtered coupled-dipole technique [19, 20] or the integration for the Green tensor [4, 21] instead of the default point-dipole formulation was shown to reduce the number of iteration. In this work, we will investigate two additional techniques for accelerating DDA simulations: starting the iterative process with an initial guess as close as possible to the actual solution and preconditioning the linear system.

2. Principle of the DDA

The discrete dipole approximation (DDA) was introduced by Purcell and Pennypacker in 1973 [22]. Since the DDA is a well known method [3, 13, 23,

4, 16], we shall outline only its main features. The object under study is represented by a cubic array of N polarizable subunits. The field at each subunit can be written:

$$\mathbf{E}(\mathbf{r}_i) = \mathbf{E}_{\text{ref}}(\mathbf{r}_i) + \sum_{j=1}^{N} \mathbf{G}(\mathbf{r}_i, \mathbf{r}_j)\alpha(\mathbf{r}_j)\mathbf{E}(\mathbf{r}_j), \tag{1}$$

where $\mathbf{E}_{\text{ref}}(\mathbf{r}_i)$ is the field at the position $\mathbf{r}_i$ in the absence of the scattering object, $\mathbf{G}$ is the dyadic tensor associated to the free space [24] or a multilayer system [25]. $\alpha(\mathbf{r}_j)$ is the polarizability of the j^{th} subunit. We use the polarizability of the Clausius-Mossotti relation with the radiative reaction term given by Draine [3]:

$$\alpha(\mathbf{r}_j) = \frac{\alpha_0(\mathbf{r}_j)}{1 - (2/3)ik_0^3\alpha_0(\mathbf{r}_j)}, \tag{2}$$

where α_0 holds the usual Clausius-Mossotti relation $\alpha_0(\mathbf{r}_j) = a^3(\varepsilon(\mathbf{r}_j)-1)/(\varepsilon(\mathbf{r}_j)+2)$. Equation (1) can be written symbolically as:

$$\mathbf{E} = \mathbf{E}_{\text{ref}} + \mathbf{A}\mathbf{D}_\alpha\mathbf{E}, \tag{3}$$

where $\mathbf{E}$, $\mathbf{E}_{\text{ref}}$ are $3N$ vectors representing the unknown field and the reference field, respectively. $\mathbf{A}$ is a $3N \times 3N$ matrix which contains the Green tensor, $\mathbf{D}_\alpha$ is a diagonal matrix of size $3N \times 3N$ which contains the polarizability. The linear system is solved iteratively, which means that we define a residual as

$$r \quad = \quad \frac{||\mathbf{E}_{\text{e}} - \mathbf{E}_{\text{ref}} - \mathbf{A}\mathbf{D}_\alpha\mathbf{E}_{\text{e}}||}{||\mathbf{E}_{\text{ref}}||}, \tag{4}$$

where $\mathbf{E}_{\text{e}}$ is the field estimated iteratively and the iterative process is terminated once $r < \eta$ where η is a prescribed tolerance. This tolerance is the parameter that quantifies the accuracy of the final field estimation. The iterative technique that is used in this article to estimate $\mathbf{E}_{\text{e}}$ is the generalized product bi-conjugate gradient solver [26] which is an alternative version of the bi-conjugate gradient stabilized solver, see Ref. [16, 17, 18] for the efficiency of different iterative methods. In the iterative process numerous MVP are required. The matrix $\mathbf{A}$ being Toeplitz, placing the dipoles on a cubic lattice: $\mathbf{i} = (i_x, i_y, i_z)$, where $i_x =$

$0, \cdots, n_x - 1$, $i_y = 0, \cdots, n_y - 1$ and $i_z = 0, \cdots, n_z - 1$ with $\mathbf{r}_i = (i_x d, i_y d, i_z d)$ and $N = n_x n_y n_z$ permits to calculate the MVP using Fast Fourier transform (FFT). Notice that objects with arbitrary shape can still be represented by a cubic lattice, by simply setting to zero the polarizations at the lattice sites lying outside the object boundaries [27]. The main advantage of DDA is that only a small portion of space containing the object is meshed and there is no need for absorbing boundary conditions, thus it is very memory efficient. Its main problem lies in the convergence of the iterative solver which can be very slow, or worse, fail, when the object is large compared to the illumination wavelength and/or is highly contrasted.

In the following, we study the influence of the initial estimate and the preconditioner on the time required for the iterative solver to converge. All the simulations were performed with the freely accessible code IFDDA which can be found in Ref. [11]. Other codes exist for DDA (see Ref. [28] appendix B for a list of Open-Source DDA), but they are all based on the same principle, so the results presented in this article are general. The Fast Fourier Transforms (FFT) are computed with the Fastest Fourier Transform in the West [29]. The calculations are parallelized (OpenMP) on a computer with 12 Intel(R) Xeon(R) CPU E5-2687W v2 at 3.40GHz.

3. Influence of the initial guess of the iterative solver on its computation time

The iterative estimation of the field inside the sample usually starts with a null field, *i.e.* $\mathbf{E}_{\text{ini}} = \mathbf{0}$. Yet, taking an initial guess that is closer to the final solution, such as the field estimated by a fast approximate method, is a better choice for diminishing the number of iterations. Hence, if one does not depart too much from the validity domain of Born approximation , *i.e.* for objects with small contrast and small size compared to the wavelength, the initial guess can be the field that would exist without the sample,

$$\mathbf{E}_{\text{ini}} \approx \mathbf{E}_{\text{ref}}. \tag{5}$$

For larger objects that mainly diffract in the forward direction, one can use the field given by the beam propagation method (BPM) [30, 31, 32] or the Rytov approximation [33]. When the sample polarizability varies slowly compared to the wavelength, one can use the Classical Scalar Approximation (CSA) [34]. This technique consists in approximating the Green's tensor appearing in the linear system Eq. (1) by the diagonal tensor $\mathbf{G} = k_0^2 \frac{e^{ik_0 r}}{r}\mathbf{I}$ where $\mathbf{I}$ is the identity tensor [35, 36, 37]. This approach yields a scalar linear system for each field component. If the incident field polarization is constant throughout the object (directed along $\hat{\mathbf{u}}$, with $\hat{\mathbf{u}}^*.\hat{\mathbf{u}} = 1$), and the field inside the object is only weakly depolarized, another scalar approximation, named uGu, can be used. By assuming that the field inside the object remains polarized along $\hat{\mathbf{u}}$, $\mathbf{E} = e\hat{\mathbf{u}}$, the uGu method transforms the vectorial linear system given by Eqs. (1) and (3) into a scalar system where the Green's tensor is replaced by the function $\hat{\mathbf{u}}^*.\mathbf{G}\hat{\mathbf{u}}$ and the unknown vector $\mathbf{E}$ by the unknown scalar component e [38]:

$$e(\mathbf{r}_i) = \mathbf{E}_{\text{ref}}(\mathbf{r}_i).\hat{\mathbf{u}}^* + \sum_{j=1}^{N} \hat{\mathbf{u}}^*.\mathbf{G}(\mathbf{r}_i, \mathbf{r}_j)\hat{\mathbf{u}}\alpha(\mathbf{r}_j)e(\mathbf{r}_j). \tag{6}$$

It was shown in Ref. [38] that the uGu method was significantly more accurate than CSA for a similar computational cost.

In the following, we consider objects that are illuminated by a plane wave so that the incident field inside the sample is directed along $\hat{\mathbf{u}}$. We study the computation time for solving the linear system Eq. (3) when the initial guess is the field given by Born, Rytov, BPM, CSA or uGu method. Note that while the field given by the Born, Rytov and BPM techniques is obtained almost instantly, that given by the CSA and uGu requires the solving of a linear system of size $N \times N$. The latter is performed with the same conjugate gradient technique as that used for solving the vectorial system, Eq. (3), except that the stopping criterion is different, η_s for the scalar problem, η for the vectorial one. Obviously, the time required for calculating the initial guess is taken into account when estimating the global computation time.

The first object under study consists of a cube of relative permittivity $\varepsilon = 1.6$

and increasing side a, the mesh-size being set to $d = \lambda/10$. The iterative solving of the vectorial system Eq. (3) was conducted with a stopping criterion $\eta = 10^{-4}$ while that of the scalar CSA or uGu systems was $\eta_s = 0.01$. We plot in Fig. 1(b) the computation time of the different methods divided by the time obtained when uGu provides the initial guess, as a function of the cube size, from $a = 10\lambda$ to $a = 20\lambda$. It is seen that the uGu field is the best initial guess for reducing the computation time, especially for large objects. Note that decreasing the computation time for small objects, $a \leq \lambda$, is of minor interest, as the calculation is very quick. For $a = 20\lambda$, using the uGu field as initial guess yields a time gain of at least 25% compared to all the other approaches. To complete this analysis, we plot in Fig. 1(a) the number of MVPs required by the iterative solver to converge depending on the initial guess. As expected, starting with the uGu field permits a significant decrease in the number of MVP. This result shows the interest of providing an accurate initial guess for starting the iterative inversion scheme, even though the latter also requires the solving of a linear system. At this point, it is worth studying the influence of the stopping criterion η_s on the time gain. Obviously, large η_s diminishes the calculation time of the scalar field but increases its inaccuracy while small η_s improves the scalar field estimation but increases its computation time. Comparing the number of MVPs required for the solving of Eq. (3) when the uGu field is estimated with η_s equal to $0.1, 0.01, 0.001$ shows that the best compromise is $\eta_s = 0.01$, see Fig. 2.

In a second example, we consider the same object deposited on a glass substrate of refractive index 1.5 and illuminated under normal incidence from the glass side or under total internal reflection (classic microscopy configuration [39]). Few approximate methods are able to simulate the field inside an object above a substrate and we could only resort to the null, Born and uGu fields for providing the initial guess. Figures 3 and 4 show the normalized computation time together with the number of MVPs required for solving the linear system with these different initial guesses. They confirm that, despite the computation cost of its calculation, the uGu field is the best option for initializing

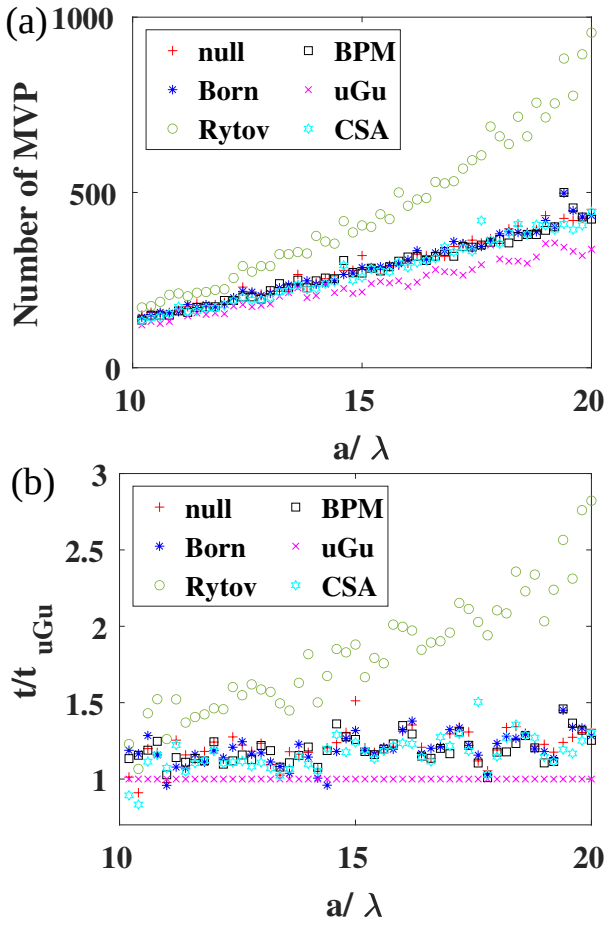

Figure 1: Cuboid of side $(a \times a \times a/2)$ and relative permittivity $\varepsilon = 1.6$ with a meshsize $d = \lambda/10$. (a) Number of MVP versus a for the different initial guess. (b) Time of computation normalized to the time used by the scalar approximation uGu.

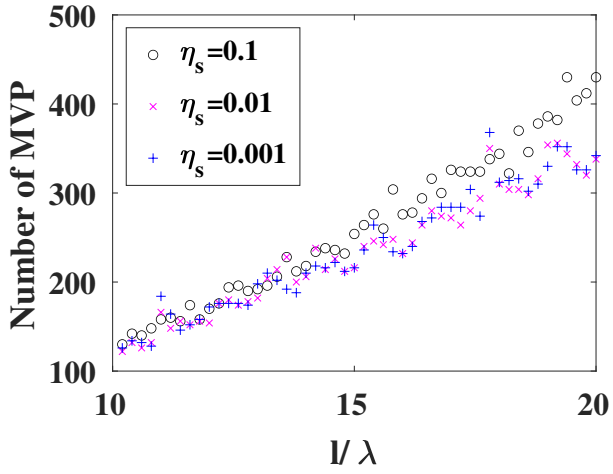

Figure 2: Number of MVPs required by the iterative solving of Eq. (3) when the initial guess is the uGu field estimated using various stopping criteria η_s. The object is a cuboid of increasing size ($a \times a \times a/2$) and relative permittivity $\varepsilon = 1.6$ with a mesh-size $d = \lambda/10$.

the iterative solver as it yields a time gain up to 20% under normal incidence and 50% under total internal reflection.

In an attempt to ameliorate further the initial guess, we implemented the Hagedüs trick which consists in optimizing a scalar so that the initial field times this scalar minimizes the error on the linear system [40]. Unfortunately, except when taking the Rytov field as initial guess, this technique did not yield any improvement (and even though faster, the Rytov solution remained the slowest).

Figure 5 studies the influence of the relative permittivity ε for a cuboid of size $(a \times a \times a/2)$ with $a/\lambda = 20$ on the computation time. Taking the uGu solution as initial guess permits a significant reduction of the computation time, up to 40%, as long as the cuboid permittivity remains smaller than 2. Indeed, cuboid with permittivities higher than 2 tend to depolarize significantly the incident beam and the field obtained with the uGu approximation is too far from the actual one [38] to be useful. Note that very recently, Inzhevatkin and Yurkin show that the approximation of WKB (Wentzel Kramers Brillouin) as an initial estimate

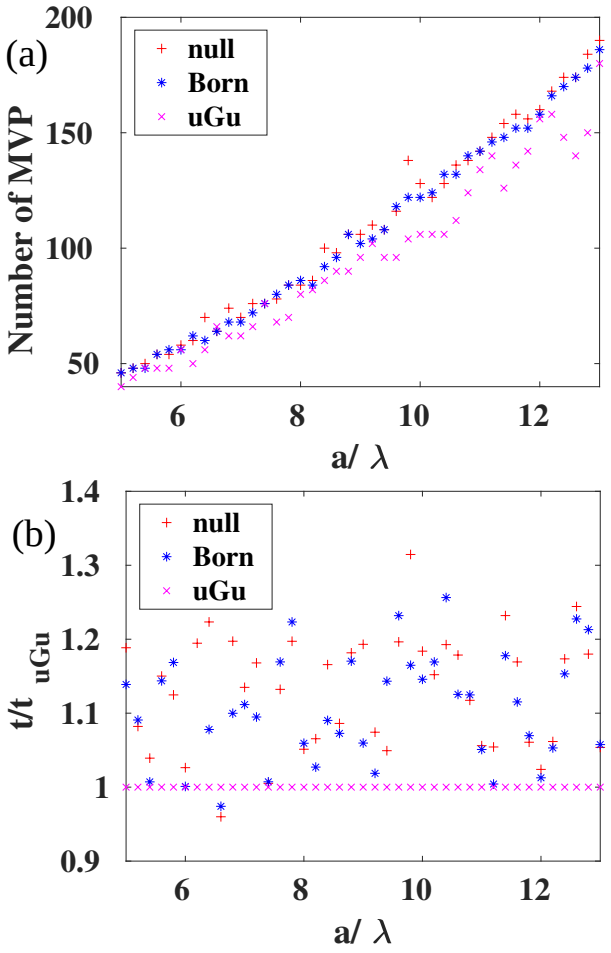

Figure 3: Cuboid of side $(a \times a \times a/2)$ and relative permittivity $\varepsilon = 1.6$ with a meshsize $d = \lambda/10$ deposited on a glass substrate. (a) Number of MVP versus a for the three different initial guess: Born, null and uGu. (b) Computation time using the Born or null field as initial guess divided by the time obtained when the scalar approximation uGu is the initial guess.

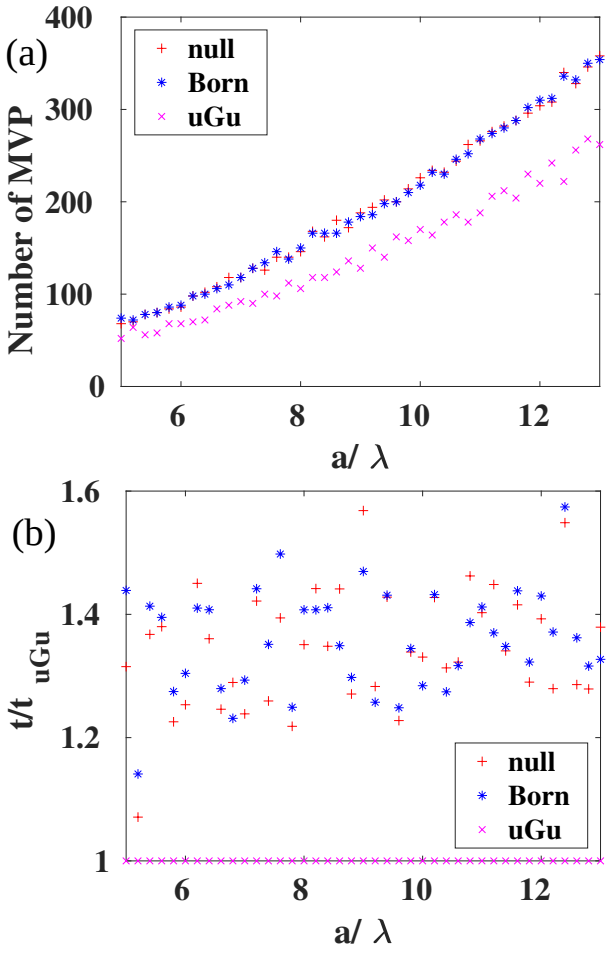

Figure 4: Same legend as in Fig. 3, but the illumination is done in total internal reflection.

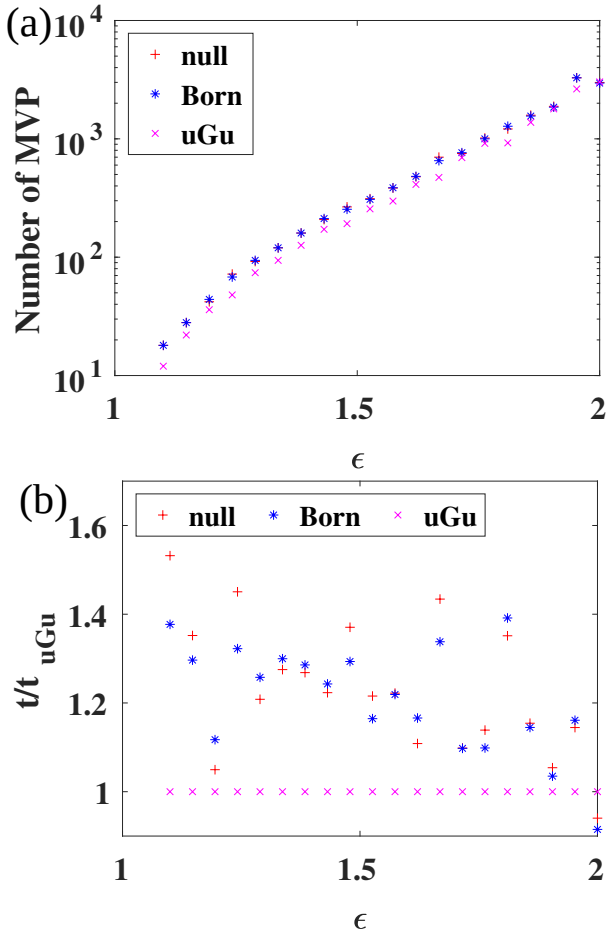

Figure 5: Cuboid of side $(a \times a \times a/2)$ with $a/\lambda = 20$ with a meshsize $d = \lambda/10$ in homogeneous space. (a) Number of MVP versus ε for the three different initial guess: Born, null and uGu. (b) Computation time using the Born or null field as initial guess divided by the time obtained when the scalar approximation uGu is the initial guess.

could also reduce the number of iterations for objects with low permittivity ($\varepsilon < 1.3$) [41]. This confirms the fact that DDA is mainly suitable for optically soft particles, when the particles become larger than wavelength of illumination and that other techniques should be preferred when the permittivity exceeds 2.5 [42]. In this case, a possible solution for speeding up the DDA consists in preconditioning the matrix of interaction to change its spectrum [43]. This is what we will study in the next section.

4. Influence of the preconditioner on the computation time

In this section, we investigate the interest of preconditioning the linear system for ameliorating the convergence of the iterative solver. This study is conducted for both the vectorial and scalar linear systems.

4.1. Construction of the preconditioner

We first rewrite the linear system to be solved Eq. (3) in the form,

$$\overline{\mathbf{A}}\mathbf{E} = \mathbf{E}_{\text{ref}}, \tag{7}$$

where $\overline{\mathbf{A}} = \mathbf{I} - \mathbf{A}\mathbf{D}_\alpha$. Instead of solving the original linear system above, one may consider the right preconditioned system,

$$\overline{\mathbf{A}}\mathbf{P}^{-1}(\mathbf{P}\mathbf{E}) = \mathbf{E}_{\text{ref}}, \tag{8}$$

and solve $\overline{\mathbf{A}}\mathbf{P}^{-1}\mathbf{y} = \mathbf{E}_{\text{ref}}$ with $\mathbf{E} = \mathbf{P}^{-1}\mathbf{y}$. Alternatively, one may solve the left preconditioned system

$$\mathbf{P}^{-1}\overline{\mathbf{A}}\mathbf{E} = \mathbf{P}^{-1}\mathbf{E}_{\text{ref}}. \tag{9}$$

Both systems give the same solution as the original system as long as the preconditioner matrix $\mathbf{P}$ is non singular. Recently, Groth *et al.* [44] proposed a left preconditioning strategy based on the two level circulant preconditioner of Chan and Olkin [45] for the DDA in the homogeneous configuration. The idea consists in finding a two level circulant matrix, the closest possible to $\overline{\mathbf{A}}$, which

can be easily invertible. To this aim, they consider a configuration in which the object polarizability, α, is assumed to be constant over the cubic lattice. For inhomogeneous samples, α is the average of the polarizability over the cubic lattice which corresponds to the Maxwell Garnett 'homogenized' medium. When α is a constant, the matrix $\overline{\mathbf{A}}$ can be written, in the general case of an object placed in a multilayered environment, as a block Toeplitz matrix (along i_y) of Toeplitz sub-matrices (along i_x). To obtain an easily invertible matrix, $\overline{\mathbf{A}}$ is approximated by a circulant matrix along these two dimensions. More precisely, the preconditioning matrix reads

$$\begin{aligned} P_{i_x,i_y,i_{z1},i_{z2}} &= \delta(\mathbf{i}-\mathbf{0}) - \alpha\frac{(n_y-i_y)}{n_y}\frac{(n_x-i_x)}{n_x}g_{i_x,i_y,i_{z1},i_{z2}} \\ &- \alpha\frac{(n_y-i_y)}{n_y}\frac{i_x}{n_x}g_{-(n_x-i_x),i_y,i_{z1},i_{z2}} \\ &- \frac{i_y}{n_y}\frac{(n_x-i_x)}{n_x}g_{i_x,-(n_y-i_y),i_{z1},i_{z2}} \\ &- \frac{i_y}{n_y}\frac{i_x}{n_x}g_{-(n_x-i_x),-(n_y-i_y),i_{z1},i_{z2}}, \end{aligned} \tag{10}$$

where g is one of the nine components of the dyadic Green's tensor and δ is the Dirac's delta function.

To obtain $\mathbf{P}^{-1}$, each 2-level circulant-block matrix $\mathbf{P}(i_{z1}, i_{z2})$ of size $n_x n_y$ is diagonalized using 2D-FFTs [46]. Then, the 9 components of the Green tensor are gathered to create a diagonal block matrix with $n_x n_y$ blocks of size $3n_z \times 3n_z$. The inversion of each block is done thanks to a LU method. For the LU method we use the zgetrs routine from LAPACK [47]. More details are given in Ref. [44]. Once $\mathbf{P}$ and $\mathbf{P}^{-1}$ are calculated, one solves iteratively the left or right conditioned linear system.

4.2. Testing the preconditioner for an object in homogeneous space

To check the interest of the preconditioning, we first consider a homogeneous cuboid of side $a = b = 10\lambda$, relative permittivity $\varepsilon = 2$, with an increasing height l_z from $\lambda/5$ to 10λ. The tolerance of the iterative method is fixed to $\eta = 10^{-4}$ and a null field is taken as initial guess. Figure 6(a) shows the

number of MVPs required by the iterative solver as a function of l_z without preconditioner (NP), left preconditioner (LP) and right preconditioner (RP). It is clear that preconditioning reduces the number of MVPs, particularly for $l_z < a/2$, see Fig. 6(a), the right preconditioner being slightly better than the left. Yet, since preconditioning requires additional computation time, we also plot the computation time of the three methods (NP, RP, LP) divided by the computation time without preconditioner (NP), see Fig. 6(b). We observe that for $l_z < a/2$ the preconditioning is efficient and time saving (except when the object is represented by two layers of dipoles). As the object becomes thicker the time gain depends on the number n_z of layers. Indeed, when the prime factors

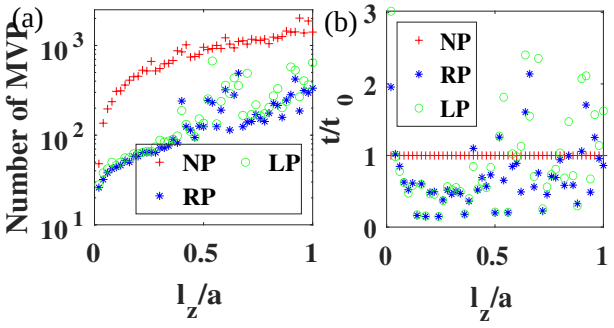

Figure 6: Solving the diffraction by a cuboid of width $a = b = 10\lambda$ and increasing height, l_z/a (a) Number of MVPs required by the iterative solver in log scale with no preconditioner (NP), left preconditioner (LP) and right preconditioner (RP). (b) Computation time required by the solving of the left or right preconditioned linear system divided by the time without preconditioner (NP).

decomposing n_z are small, the 3D FFTs that are used for solving the linear system in absence of preconditioner (we recall that, in homogeneous space, $\mathbf{A}$ is block Toeplitz along the three directions (i_x, i_y, i_z) while $\mathbf{P}$ is circulant only along i_x and i_y) are very fast and the preconditioning is useless. On the contrary, when the prime factors of n_z are large, the 3D FFTs are less efficient and preconditioning the system is interesting. It allows a faster convergence of the iterative solver, as shown in Fig. 7(a) where the residue of the iterative

solver is plotted versus the number of MVPs for NP, LP and RP for a cuboid with $a = b = 10\lambda$, $l_z = a/5$ and relative permittivity 2.

In Fig. 7(b) we investigate the interest of the preconditioning when the permittivity of the cuboid of Figure 7(a) is increased. It is shown that preconditioning is interesting when the permittivity contrast is moderate (ε between 1.2 to 2.3). For weak permittivities, NP requires few MVPs and preconditioning is useless, for strong permittivities (above 2.5) the convergence of the preconditioned system appears less stable than the non-reconditioned one.

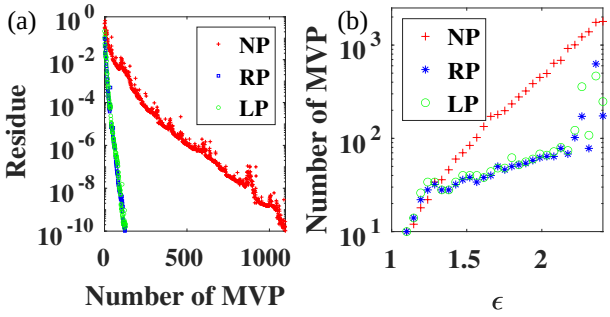

Figure 7: Cuboid of side $a = b = 10\lambda$ and $l_z = a/5$. (a) Evolution of the residue in log scale as a function of the number of MVPs of the iterative method for no preconditioning (NP), with a right (RP) and left preconditioner (LP) for $\varepsilon = 2$. (b) Number of MVPs in log scale versus the permittivity for $\eta = 10^{-4}$.

We now consider an inhomogeneous sample corresponding to a cuboid of side $a = b = 10\lambda$ and $l_z = a/5$ and random relative permittivity with Gaussian probability density of mean $\varepsilon_{\rm bg}$ and variance σ^2, with a Gaussian correlation function [48], defined by $\langle \varepsilon(\mathbf{r}), \varepsilon(\mathbf{r'}) \rangle = \varepsilon_{\rm bg}^2 + \sigma^2 \exp\left(-\frac{\|\mathbf{r}-\mathbf{r'}\|^2}{l_c^2}\right)$. Figure 8 plots the number of MVP and normalized time versus σ for $l_c = \lambda$ and $\varepsilon_{\rm bg} = 2$.

We observe that the preconditioning reduces the computation time by a factor of two when the inhomogeneity is moderate but is clearly inappropriate when the inhomogeneity is strong. We tried to improve its performance by extending to the two-dimensional case the optimal circulant preconditioner presented in [49] for one-dimensional problems. Unfortunately, this modification did not

yield any amelioration.

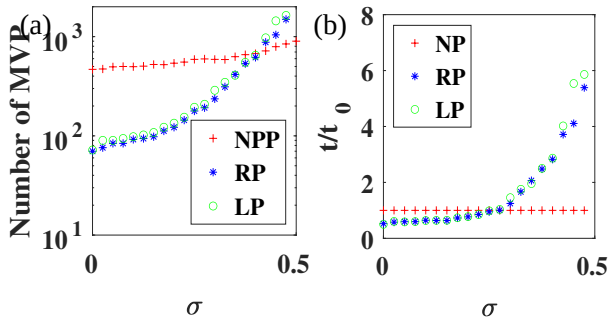

Figure 8: Inhomogeneous cuboid of size $(10\lambda \times 10\lambda \times 2\lambda)$ with $\varepsilon_{\text{bg}} = 2$ and $l_c = \lambda$. (a) Number of MVPs versus σ for no preconditioning (NP) with a right (RP) and left preconditioner (LP). (b) Time of computation with left (LP) and right (RP) preconditioner divided by the computation time obtained with no preconditioner (NP).

4.3. Testing the preconditioner for the scalar approximation

Recently we have introduced a scalar approximation that yields accelerate estimates of the field calculations with good accuracy for permittivities typically less than two [38]. Under this approximation, the 2-level Chan preconditioner is easy to implement as the Green's function has only one component. It becomes a diagonal block matrix made of $n_x n_y$ blocks of size $n_z \times n_z$ the inversion of which is fast. We study the time gain brought by the preconditioning on the same cuboid of increasing height as before. Figure 9 shows that whatever the height, the number of MVPs and the computation time is drastically reduced by the preconditioner. Another study conducted with the inhomogeneous cuboid revealed the same behavior as with the full vectorial problem : the preconditioner is efficient for moderate inhomogeneity.

4.4. Testing the preconditioner for an object in presence of a substrate

For an object in the presence of a multilayer, the **A** matrix in Eq. (3) is only 2D Toeplitz in the x and y direction. This means that, except when the

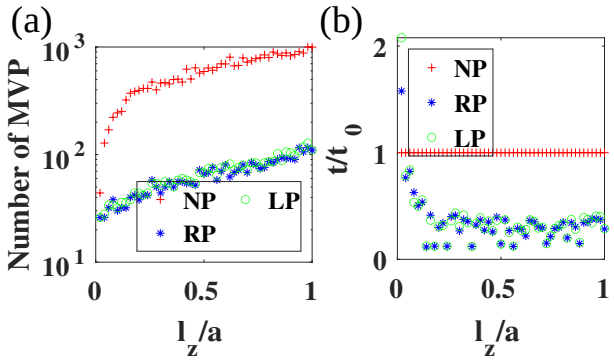

Figure 9: Same as Fig. 6 but the vectorial linear system was replaced by the scalar linear system using the uGu approximation.

object is in the substrate or superstrate [51], the matrix vectors product in the z-direction can not be done by FFT [50]. This configuration seems thus better adapted to the use of the two-level circulating preconditioner. To test its performance, we study the same configurations as previously but the cuboids are now deposited on a substrate of permittivity 2.25 and illuminated under normal incidence. We observe in Figs. 10(a) and 10(b) that, in this case, the preconditioner reduces the number of MVPs and allows a significant time gain whatever the height of the cuboid.

Similarly, when the permittivity of the cuboid is moderately increased, Fig. 11, the time gain brought by the preconditioner is important and can reach a factor of four. Yet, similarly to the homogeneous case, the preconditioner renders the iterative solver unstable for permittivity above $\varepsilon = 2.5$.

5. Conclusion

The new formulation of the scalar approximation that we have introduced in [38] is an excellent initial estimate for the iterative method that solves the DDA linear equation system. It reduces the computation time by 50% for moderately contrasted objects (permittivity below 2) such as those encountered

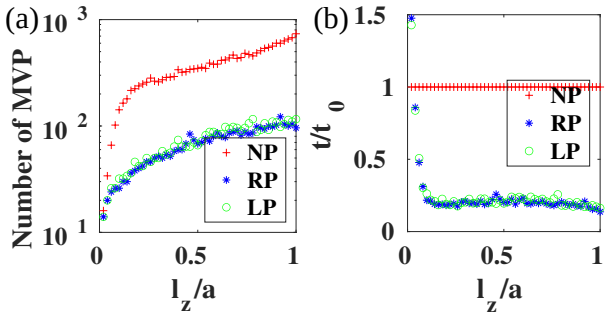

Figure 10: Same as Fig. 6 but the cuboid is placed on a glass substrate located at $z = 0$ and the illumination is done along the z axis from the glass side.

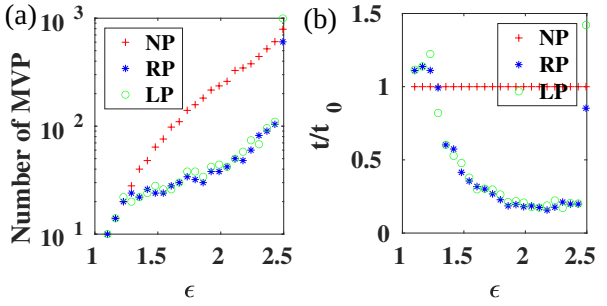

Figure 11: Cuboid of size $(10\lambda \times 10\lambda \times 2\lambda)$ placed on a glass substrate located at $z = 0$ and the illumination is done along the z axis from the glass side. (a) Number of MVPs in log scale as a function of ε with no preconditioner (NP), left preconditioner (LP) and right preconditioner (RP). (b) Time of computation normalized with respect to the computation time without preconditioner (NP) for the right and left preconditioning.

in biological applications. We observed that this initial estimate was particularly efficient when the illumination was inhomogeneous, as shown for evanescent illumination.

In addition, we have studied the interest of the 2-level circulating left preconditioner introduced by Groth *et al.* [44] for further diminishing the computation time. Groth *et al.* showed that this preconditioner was efficient for homogeneous flat object. We extended it to inhomogeneous objects and showed its general interest for speeding up the solving of Maxwell equation under the scalar approximation. In the vectorial case, its interest was more limited and concerned weakly inhomogeneous objects with moderate permittivity deposited on a substrate.

References

[1] F. M. Kahnert, Numerical methods in electromagnetic scattering theory, J. Quant. Spect. Rad. Transf. 79-80 (2003) 775–824.

[2] B. Gallinet, J. Butet, O. J. F. Martin, Numerical methods for nanophotonics: standard problems and future challenges, Laser & Photonics Reviews 9 (6) (2015) 577–603.

[3] B. T. Draine, The discrete-dipole approximation and its application to interstellar graphite grains, Astrophys. J. 333 (1988) 848–872.

[4] P. C. Chaumet, A. Sentenac, A. Rahmani, Coupled dipole method for scatterers with large permittivity, Phys. Rev. E 70 (2004) 036606–6.

[5] F. Moreno, R. Vilaplana, O. Muñoz, A. Molina, D. Guirado, The scattering matrix for size distributions of irregular particles: An application to an olivine sample, Journal of Quantitative Spectroscopy and Radiative Transfer 100 (1) (2006) 277–287, vIII Conference on Electromagnetic and Light Scattering by Nonspherical Particles.

[6] M. A. Yurkin, D. de Kanter, A. G. Hoekstra, Accuracy of the discrete dipole approximation for simulation of optical properties of gold nanoparticles, Journal of Nanophotonics 4 (1) (2010) 1 – 15 – 15.

[7] M. A. Yurkin, A. G. Hoekstra, The discrete dipole approximation: An overview and recent developments, J. Quant. Spect. Rad. Transf. 106 (2007) 558–589.

[8] M. A. Yurkin, V. P. Maltsev, A. G. Hoekstra, The discrete dipole approximation for simulation of light scattering by particles much larger than the wavelength, J. Quant. Spect. Rad. Transf. 106 (2007) 546–557.

[9] B. T. Draine, P. J. Flatau, Discrete-dipole approximation for periodic targets: theory and tests, J. Opt. Soc. Am. A 25 (11) (2008) 2693–2703.

[10] M. A. Yurkin, A. G. Hoekstra, The discrete-dipole-approximation code adda: Capabilities and known lim itations, J. Quant. Spect. Rad. Transf. 112 (13) (2011) 2234 – 2247.

[11] P. C. Chaumet, D. Sentenac, G. Maire, M. Rasedujjaman, T. Zhang, A. Sentenac, Ifdda, an easy-to-use code for simulating the field scattered by 3d inhomogeneous objects in a stratified medium: tutorial, J. Opt. Soc. Am. A 38 (12) (2021) 1841–1852.

[12] P. C. Chaumet, T. Zhang, A. Sentenac, Fast far-field calculation in the discrete dipole approximation, Journal of Quantitative Spectroscopy and Radiative Transfer 165 (2015) 88 – 92.

[13] P. J. Flatau, G. L. Stephens, B. T. Draine, Light scattering by rectangular solids in the discrete-dipole approximation: a new algorithm exploiting the Block-Toeplitz structure, J. Opt. Soc. Am. A 7 (1990) 593–600.

[14] R. D. Da Cunha, T. Hopkins, The Parallel Iterative Methods (PIM) package for the solution of systems of linear equations on parallel computers, Appl. Numer. Math. 19 (1995) 33–50.

[15] P. J. Flatau, Improvements in the discrete-dipole approximation method of computing scattering and absorption, Opt. Lett. 22 (1997) 1205–1207.

[16] P. C. Chaumet, A. Rahmani, Efficient iterative solution of the discrete dipole approximation for magneto-dielectric scatterers , Opt. Lett. 34 (2009) 917–919.

[17] K. Skorupski, Using the dda (discrete dipole approximation) method in determining the extinction cross sectionof black carbon, Metrology and Measurement Systems 22 (2015) 153–164.

[18] M. A. Yurkin, Performance of iterative solvers in the discrete dipole approximation, in: 2016 URSI International Symposium on Electromagnetic Theory (EMTS), 2016, pp. 488–491.

[19] N. Piller, O. Martin, Increasing the performance of the coupled-dipole approximation: a spectral approach, IEEE Transactions on Antennas and Propagation 46 (8) (1998) 1126–1137.

[20] M. A. Yurkin, M. Min, A. G. Hoekstra, Application of the discrete dipole approximation to very large refractive indices: Filtered coupled dipoles revived, Phys. Rev. E 82 (2010) 036703.

[21] D. A. Smunev, P. C. Chaumet, M. A. Yurkin, Rectangular dipoles in the discrete dipole approximation, J. Quant. Spect. Rad. Transf. 156 (0) (2015) 67 – 79.

[22] E. M. Purcell, C. R. Pennypacker, Scattering and absorption of light by nonspherical dielectric grains, Astrophys. J. 186 (1973) 705–714.

[23] B. T. Draine, J. Goodman, Beyond Clausius-Mossotti: Wave Propagation on a Polarizable Point Lattice and the Discrete Dipole Approximation, Astrophys. J. 405 (1993) 685–697.

[24] J. D. Jackson, Classical Electrodynamics, 2nd Edition, Wiley, 1975.

[25] M. Paulus, P. Gay-Balmaz, O. J. F. Martin, Accurate and efficient computation of the Green's tensor for stratified media, Phys. Rev. E 62 (4) (2000) 5797–5807.

[26] S.-L. Zhang, Gpbi-cg: Generalized product-type methods based on bi-cg for solving nonsymmetric linear systems, SIAM Journal on Scientific Computing 18 (2) (1997) 537–551.

[27] J. J. Goodman, P. J. Flatau, Application of fast-fourier-transform techniques to the discrete-dipole approximation, Opt. Lett. 16 (2002) 1198–1200.

[28] P. C. Chaumet, The discrete dipole approximation: A review, Mathematics 10 (17) (2022).

[29] M. Frigo, S. G. Johnson, The design and implementation of FFTW3, Proceedings of the IEEE 93 (2) (2005) 216–231, special issue on "Program Generation, Optimization, and Plat form Adaptation".

[30] U. S. Kamilov, I. N. Papadopoulos, M. H. Shoreh, A. Goy, C. Vonesch, M. Unser, D. Psaltis, Learning approach to optical tomography, Optica 2 (6) (2015) 517–522.

[31] U. S. Kamilov, I. N. Papadopoulos, M. H. Shoreh, A. Goy, C. Vonesch, M. Unser, D. Psaltis, Optical tomographic image reconstruction based on beam propagation and sparse regularization, IEEE Transactions on Computational Imaging 2 (1) (2016) 59–70.

[32] P. C. Chaumet, A. Sentenac, T. Zhang, Reflection and transmission by large inhomogeneous media. validity of b orn, rytov and beam propagation methods, Journal of Quantitative Spectroscopy and Radiative Transfer 243 (2020) 106816.

[33] R. Carminati, Phase properties of the optical near field, Phys. Rev. E 55 (1997) R4901–R4904.

[34] M. Born, E. Wolf, Principles of Optics, Pergamon, 1959.

[35] P. S. Carney, J. C. Schotland, Inverse scattering for near-field microscopy, Appl. Phys. Lett. 77 (18) (2000) 2798–2800.

[36] P. S. Carney, J. C. Schotland, Theory of total-internal-reflection tomography, J. Opt. Soc. Am. A 20 (2003) 542–547.

[37] P. S. Carney, J. C. Schotland, Three-dimensional total-internal reflection microscopy, Opt. Lett. 26 (14) (2001) 1072–1074.

[38] P. C. Chaumet, G. Maire, A. Sentenac, Scalar approximation of maxwell equations: derivation and accuracy, J. Opt. Soc. Am. A 39 (8) (2022) 1462–1467.

[39] P. C. Chaumet, K. Belkebir, A. Sentenac, Superresolution of three-dimensional optical imaging by use of evanescent waves, Opt. Lett. 29 (2004) 2740–2742.

[40] Z. Strakoš, J. Liesen, On numerical stability in large scale linear algebraic computations, Z. Angew. Math. Mech. 85 (5) (2005) 307–325.

[41] K. G. Inzhevatkin, M. A. Yurkin, Uniform-over-size approximation of the internal fields for scatterers with low refractive-index contrast, Journal of Quantitative Spectroscopy and Radiative Transfer 277 (2022) 107965.

[42] C. Liu, L. Bi, and R. Lee Panetta, P. Yang and M. A. Yurkin, Comparison between the pseudo-spectral time domain method and the discrete dipole approximation for light scattering simulations, Opt. Express 20 (15) (2012) 16763–16776.

[43] X.-G. Lv, T.-Z. Huang, Z.-G. Ren, A modified t. chan's preconditioner for toeplitz systems, Computers & Mathematics with Applications 58 (4) (2009) 693–699.

[44] S. P. Groth, A. G. Polimeridis, J. K. White, Accelerating the discrete dipole approximation via circulant preconditioning, Journal of Quantitative Spectroscopy and Radiative Transfer 240 (2020) 106689.

[45] T. F. Chan, J. A. Olkin, Circulant preconditioners for toeplitz-block matrices, Numer. Algor. 6 (1994) 89–101.

[46] R. H. Chan, J. G. Nagy, R. J. Plemmons, Fft-based preconditioners for toeplitz-block least squares problems, SIAM Journal on Scientific and Statistical Computing 30 (6) (1992) 1740–1768.

[47] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, D. Sorensen, LAPACK Users' Guide, 3rd Edition, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1999.

[48] T. Zhang, P. C. Chaumet, E. Mudry, A. Sentenac, K. Belkebir, Electromagnetic wave imaging of targets buried in a cluttered medium u sing a hybrid inversion-dort method, Inverse Probl. 28 (12) (2012) 125008.

[49] R. F. Remis, Circulant preconditioners for domain integral equations in electromagnetics, Electromagnetics in advanced applications (ICEAA), international conference on. IEEE 85 (2012) 337–340.

[50] R. Schmehl, B. M. Nebeker, E. D. Hirleman, Discrete-dipole approximation for scattering by features on surfaces by means of a two-dimensional fast fourier transform technique, J. Opt. Soc. Am. A 14 (11) (1997) 3026–3036.

[51] M. A. Yurkin, M. Huntemann, Rigorous and fast discrete dipole approximation for particles near a plane interface, J. Phys. Chem. C 119 (52) (2015) 29088–29094.