

The validation of the parallel three-dimensional solver for analysis of optical plasmonic bi-periodic multilayer nanostructures

Xingjie Ni · Zhengtong Liu · Alexandra Boltasseva · Alexander V. Kildishev

Received: 30 December 2009 / Accepted: 1 June 2010 / Published online: 30 June 2010
© Springer-Verlag 2010

Abstract Fundamentals of the three-dimensional spatial harmonic analysis (SHA) approach are reviewed, and the advantages of a fast-converging formulation versus the initial SHA formulation are emphasized with examples using periodic plasmonic nanostructures. First, two independent parallel versions of both formulations are implemented using the scattering matrix algorithm for multilayer cascading. Then, by comparing the results from both formulations, it is shown that choosing an advanced fast-converging scheme could be essential for accurate and efficient modeling of plasmonic structures. Important obstacles to the fast parallel implementation of this approach are also revealed. The results of test simulations are validated using the data obtained from a commercial finite-element method (FEM) simulations and from the experimental characterization of fabricated samples.

1 Introduction

The spatial harmonic analysis (SHA) method, also known as the Fourier modal method (FMM) or rigorous coupled

wave analysis (RCWA), is one of the most versatile methods for analyzing the diffraction of electromagnetic waves by periodic structures. Our SHA approach is a non-iterative and mesh-free technique for obtaining the exact solution of Maxwell's equations using the Floque–Bloch formalism. The accuracy of the solution depends solely on the number of modes in the spatial harmonic expansion of the field.

2D formulations of SHA and related methods were initially done in the 1950s, and since then, were extended and generalized by many researchers [1–13]. Later, it has been shown that the SHA approach can be used to analyze three dimensional bi-periodic structures [14–16]. In 1996, a fast-convergence formulation was proposed to solve the slow convergence problem for TM polarization (magnetic-field vector perpendicular to the grating) by Lalanne and Morris [17], and Granet and Guizal [18], and the mathematical foundation of this formulation was given in [19], Sect. 3.8. Theoretical aspects of this approach for 3D bi-periodic structures has been developed by Li [20]. The recursive matrix algorithm for cascading multiple layers was also studied intensively [21–25]. Nowadays, the SHA method is used for analyzing sub-wavelength-periodic structures, such as nanostructured metamaterials [26–29].

In this paper, the formulation of the SHA method for three-dimensional structures (bi-periodic multilayer structures) is presented based on previous works. The scattering matrix algorithm is used to cascade multiple layers. Numerical tests are conducted based on a nanoantenna sample and the results are compared with those obtained from experiments and finite-element method simulations. The convergence of the initial formulation and the fast-convergence formulation is studied and compared.

X. Ni · Z. Liu · A. Boltasseva · A.V. Kildishev (✉)
Birck Nanotechnology Center, School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA
e-mail: Kildishev@purdue.edu

A. Boltasseva
DTU Fotonik, Technical University of Denmark, 2800 Kongens Lyngby, Denmark

A. Boltasseva
Erlangen Graduate School in Advanced Optical Technologies—SAOT, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91052 Erlangen, Germany

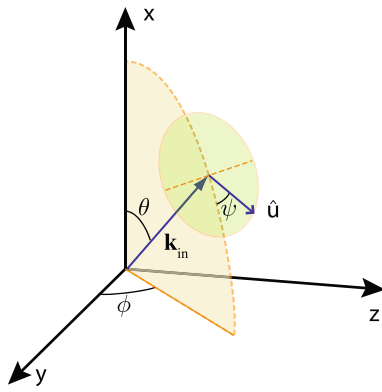


Fig. 1 Diagram of the k vector of a linear polarized incident light

2 Theory

2.1 Plane wave incidence

In order to simulate the electromagnetic fields in our model system, we must consider the mathematics used to describe the fields. We assume a plane-wave incidence and denote the k vector of the incident wave as \vec{k}_{in} as shown in Fig. 1. Using $\vec{k}_{in} = k_{in,x}\hat{x} + k_{in,y}\hat{y} + k_{in,z}\hat{z}$, where

$$k_{in,x} = n_{sup} \cos \theta, \tag{1}$$

$$k_{in,y} = n_{sup} \sin \theta \cos \phi, \tag{2}$$

$$k_{in,z} = n_{sup} \sin \theta \sin \phi, \tag{3}$$

and n_{sup} is the refractive index at the source (input) side, the electric field incident on the structure can be defined as $\vec{E}_{in} = \vec{E}_{in,0} \exp(jk_0\vec{k}_{in} \cdot \vec{r})$. The time dependence $\exp(-j\omega t)$ is assumed throughout this paper. The incident magnetic field is obtained through $\vec{H}_{in} = j(\mu\omega)^{-1} \nabla \times \vec{E}_0$ or $\vec{H}_{in} = -(\eta_0\eta_r)^{-1} \vec{k}_{in} \times \vec{E}_{in,0} \exp(jk_0\vec{k}_{in} \cdot \vec{r})$, where $\eta_0 = \sqrt{\mu_0/\epsilon_0}$, is the impedance of free space and $\eta_r = \sqrt{\mu_r/\epsilon_r}$, is the relative impedance. We also assume that the incident plane wave is linearly polarized. The polarization vector \hat{u} , which denotes the direction of the incident magnetic field in this paper, can be written as $\hat{u} = u_x\hat{x} + u_y\hat{y} + u_z\hat{z}$, where

$$u_x = -\cos \psi \sin \theta, \tag{4}$$

$$u_y = \cos \psi \cos \theta \cos \phi - \sin \psi \sin \phi, \tag{5}$$

$$u_z = \cos \psi \cos \theta \sin \phi + \sin \psi \cos \phi. \tag{6}$$

2.2 Eigenvalue problem

We start with Maxwell’s curl equations $\nabla \times \vec{E} = j\omega\vec{B}$ and $\nabla \times \vec{H} = -j\omega\vec{D}$, and since $\nabla \cdot \vec{H} = 0$, i.e. $\nabla \times \nabla \times \vec{H} = -\nabla^2\vec{H}$, we arrive at

$$-\epsilon_r^{-1}\nabla^2\vec{H} = -\nabla\epsilon^{-1} \times \nabla \times \vec{H} + k_0^2\vec{H}. \tag{7}$$

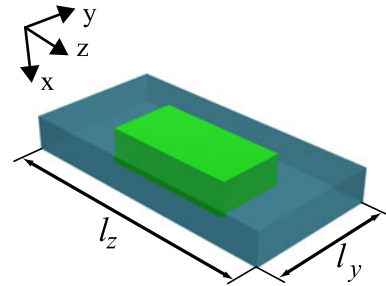


Fig. 2 Sketch of an example three-dimensional periodic unit cell

We note that the structure under analysis has permittivity periodically modulated along the y and z directions and invariant along the x direction. The periods along y and z directions are l_y and l_z , respectively. An example unit cell of such structure is shown in Fig. 2. Thus, each field component can be expressed as

$$H_i = \tilde{v}_i U c_i, \quad i = x, y, z, \tag{8}$$

where c_i is a vector of relative amplitude coefficients. Namely, each field component is written as the product of a periodic part $\tilde{v}_i = \tilde{v}_i(y, z)$ and a wavelike part

$$U = \text{diag}[\exp(jk_0k_x^{(-M_x)}x), \dots, \exp(jk_0k_x^{(M_x)}x)] \tag{9}$$

as a consequence of Bloch’s theorem, where $\text{diag}[\cdot]$ denotes the arrangement of a diagonal matrix with the elements of a vector on the main diagonal. m_x is the mode number, which ranges from $-\infty$ to ∞ in theory. However, in order to make it numerically feasible, we truncate the infinite series to only keep the modes from $-M_x$ to M_x . Note that the following discussion is based on this truncation criteria.

By denoting $\partial/\partial x \equiv \partial_x, \partial/\partial y \equiv \partial_y, \partial/\partial z \equiv \partial_z, \partial^2/\partial x^2 \equiv \partial_{xx}$ and so on, we can get the following equations from (7):

$$\begin{aligned} \epsilon_r(\partial_{xx}H_y + \partial_{yy}H_y + \partial_{yz}H_z) + k_0^2H_y \\ = \partial_z[\epsilon_r^{-1}(\partial_yH_z - \partial_zH_y)], \end{aligned} \tag{10}$$

$$\begin{aligned} \epsilon_r(\partial_{xx}H_z + \partial_{yy}H_z + \partial_{zy}H_z) + k_0^2H_z \\ = \partial_y[\epsilon_r^{-1}(\partial_zH_y - \partial_yH_z)]. \end{aligned} \tag{11}$$

We expand $\tilde{v}(y, z)$ in the 2-dimensional Fourier space and let \mathbf{v} be the basis of the space, which is expressed as

$$\mathbf{v} = [v^{(-M_y, -M_z)}, \dots, v^{(-M_y, M_z)}, v^{(-M_y+1, -M_z)}, \\ v^{(-M_y+1, -M_z+1)}, \dots, v^{(M_y, M_z)}]^T. \tag{12}$$

Each element in \mathbf{v} is expressed as

$$v^{(m_y, m_z)} = \exp[j(k_0k_y^{(m_y)}y + k_0k_y^{(m_z)}z)], \tag{13}$$

where $k_y^{(m_y)} = k_{in,y} + m_y\lambda/l_y, k_z^{(m_z)} = k_{in,z} + m_z\lambda/l_z, m_y$ and m_z are the mode numbers, which can range from

$-\infty$ to ∞ . Since here we already chose to use a truncated version for the numerical computation, m_y and m_z are from $-M_y$ to M_y and from $-M_z$ to M_z , respectively. $k_{in,y}$ and $k_{in,z}$ are defined in Sect. 2.1. The index number of n th component in vector \mathbf{v} satisfies the index equation

$$n = (m_y + M_y)(2M_z + 1) + m_z + M_z + 1. \tag{14}$$

Thus, (8) can be expressed in the following matrix form

$$H_i = \mathbf{v}^T \tilde{\mathbf{H}} \mathbf{U} \mathbf{c}_i, \quad i = x, y, z, \tag{15}$$

where $\mathbf{v} = \mathbf{v}(y, z)$ and $\mathbf{v}_i = \mathbf{v}^T \tilde{\mathbf{H}}_i$. We note that \mathbf{c}_i is redundant and can be incorporated into $\tilde{\mathbf{H}}_i$. Thus (15) can be rewritten as

$$H_i = \mathbf{v}^T \mathbf{H}_i \mathbf{v}, \quad i = x, y, z, \tag{16}$$

where \mathbf{v} is a vector consisting of the diagonal elements of matrix \mathbf{U} , and \mathbf{H}_i is a $(2M_y + 1)(2M_z + 1) \times M_x$ dimensional coefficient matrix. Hence, using matrix notation, (10) and (11) can be rewritten as

$$-\epsilon_r^{-1} \mathbf{v}^T (\mathbf{H}_y \mathbf{K}_x^2 \mathbf{v} + \mathbf{K}_y^2 \mathbf{H}_y \mathbf{v} + \mathbf{K}_y \mathbf{K}_z \mathbf{H}_z \mathbf{v}) + \mathbf{v}^T \mathbf{H}_y \mathbf{v} = k_0^{-1} \partial_z [j \epsilon_r^{-1} (\mathbf{v}^T \mathbf{K}_y \mathbf{H}_z \mathbf{v} - \mathbf{v}^T \mathbf{K}_z \mathbf{H}_y \mathbf{v})], \tag{17}$$

$$-\epsilon_r^{-1} \mathbf{v}^T (\mathbf{H}_z \mathbf{K}_x^2 \mathbf{v} + \mathbf{K}_y^2 \mathbf{H}_z \mathbf{v} + \mathbf{K}_z \mathbf{K}_y \mathbf{H}_y \mathbf{v}) + \mathbf{v}^T \mathbf{H}_z \mathbf{v} = k_0^{-1} \partial_y [j \epsilon_r^{-1} (\mathbf{v}^T \mathbf{K}_z \mathbf{H}_y \mathbf{v} - \mathbf{v}^T \mathbf{K}_y \mathbf{H}_z \mathbf{v})], \tag{18}$$

where \mathbf{K}_x , \mathbf{K}_y and \mathbf{K}_z are diagonal matrices, which have the following form:

$$\mathbf{K}_x = \text{diag}[k_x^{(1)}, k_x^{(2)}, \dots, k_x^{(M_x)}], \tag{19}$$

$$\mathbf{K}_y = \text{diag}[k_y^{(-M_y)}, k_y^{(-M_y+1)}, \dots, k_y^{(M_y)}], \tag{20}$$

$$\mathbf{K}_z = \text{diag}[\mathbf{k}_z, \mathbf{k}_z, \dots, \mathbf{k}_z]_{2M_z+1}, \tag{21}$$

where $\mathbf{k}_y^{(m_y)} = \text{diag}[k_y^{(m_y)}, k_y^{(m_y)}, \dots, k_y^{(m_y)}]_{2M_z+1}$, $\mathbf{k}_z = \text{diag}[k_z^{(-M_z)}, k_z^{(-M_z+1)}, \dots, k_y^{(M_z)}]$. By expanding ϵ_r^{-1} with respect to the basis \mathbf{v}

$$\epsilon_r^{-1} = \sum_{p,q} \tilde{\gamma}^{(p,q)} \exp \left[j \left(\frac{p\lambda}{l_x} x + \frac{q\lambda}{l_y} y \right) \right]. \tag{22}$$

Then, we build a matrix $\tilde{\mathbf{\Gamma}}$ such that

$$\mathbf{v}^T \tilde{\mathbf{\Gamma}} = \epsilon_r^{-1} \mathbf{v}^T. \tag{23}$$

By matching the coefficients of the exponentials, we can find the relationship between the element in row r column s of the matrix $\tilde{\mathbf{\Gamma}}$ and the Fourier coefficients in (22). We use the four indices m, n, p and q to determine the $(m - n, p - q)$ th Fourier coefficient, n and p ranging from $-M_y$

to M_y , where n and q range from $-M_z$ to M_z , and their relationship can be expressed using the following index equations:

$$r = (m + M_y)(2M_z + 1) + n + M_z + 1, \tag{24}$$

$$s = (p + M_y)(2M_z + 1) + q + M_z + 1. \tag{25}$$

Thus, by using (23) we can rewrite (17) and (18) as

$$\mathbf{v}^T [\mathbf{I} - (\tilde{\mathbf{\Gamma}} \mathbf{H}_y \mathbf{K}_x^2 \mathbf{H}_y^{-1} + \tilde{\mathbf{\Gamma}} \mathbf{K}_y^2 + \mathbf{K}_z \tilde{\mathbf{\Gamma}} \mathbf{K}_z)] \mathbf{H}_y \mathbf{v} = \mathbf{v}^T (\tilde{\mathbf{\Gamma}} \mathbf{K}_y \mathbf{K}_z - \mathbf{K}_z \tilde{\mathbf{\Gamma}} \mathbf{K}_y) \mathbf{H}_z \mathbf{v}, \tag{26}$$

$$\mathbf{v}^T [\mathbf{I} - (\tilde{\mathbf{\Gamma}} \mathbf{H}_z \mathbf{K}_x^2 \mathbf{H}_z^{-1} + \tilde{\mathbf{\Gamma}} \mathbf{K}_z^2 + \mathbf{K}_y \tilde{\mathbf{\Gamma}} \mathbf{K}_y)] \mathbf{H}_z \mathbf{v} = \mathbf{v}^T (\tilde{\mathbf{\Gamma}} \mathbf{K}_z \mathbf{K}_y - \mathbf{K}_y \tilde{\mathbf{\Gamma}} \mathbf{K}_z) \mathbf{H}_y \mathbf{v}, \tag{27}$$

where \mathbf{I} is the identity matrix. The coefficients of each mode should be matched, thus (26) and (27) can be rewritten as two matrix equations

$$[\mathbf{I} - (\tilde{\mathbf{\Gamma}} \mathbf{H}_y \mathbf{K}_x^2 \mathbf{H}_y^{-1} + \tilde{\mathbf{\Gamma}} \mathbf{K}_y^2 + \mathbf{K}_z \tilde{\mathbf{\Gamma}} \mathbf{K}_z)] \mathbf{H}_y = (\tilde{\mathbf{\Gamma}} \mathbf{K}_y \mathbf{K}_z - \mathbf{K}_z \tilde{\mathbf{\Gamma}} \mathbf{K}_y) \mathbf{H}_z, \tag{28}$$

$$[\mathbf{I} - (\tilde{\mathbf{\Gamma}} \mathbf{H}_z \mathbf{K}_x^2 \mathbf{H}_z^{-1} + \tilde{\mathbf{\Gamma}} \mathbf{K}_z^2 + \mathbf{K}_y \tilde{\mathbf{\Gamma}} \mathbf{K}_y)] \mathbf{H}_z = (\tilde{\mathbf{\Gamma}} \mathbf{K}_z \mathbf{K}_y - \mathbf{K}_y \tilde{\mathbf{\Gamma}} \mathbf{K}_z) \mathbf{H}_y. \tag{29}$$

By combining (28) and (29), we obtain a system

$$\begin{bmatrix} \mathbf{I} - \tilde{\mathbf{\Gamma}} \mathbf{K}_y^2 - \mathbf{K}_z \tilde{\mathbf{\Gamma}} \mathbf{K}_z & \mathbf{K}_z \tilde{\mathbf{\Gamma}} \mathbf{K}_y - \tilde{\mathbf{\Gamma}} \mathbf{K}_y \mathbf{K}_z \\ \mathbf{K}_y \tilde{\mathbf{\Gamma}} \mathbf{K}_z - \tilde{\mathbf{\Gamma}} \mathbf{K}_z \mathbf{K}_y & \mathbf{I} - \tilde{\mathbf{\Gamma}} \mathbf{K}_z^2 - \mathbf{K}_y \tilde{\mathbf{\Gamma}} \mathbf{K}_y \end{bmatrix} \begin{bmatrix} \mathbf{H}_y \\ \mathbf{H}_z \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{\Gamma}} & \mathbf{O} \\ \mathbf{O} & \tilde{\mathbf{\Gamma}} \end{bmatrix} \begin{bmatrix} \mathbf{H}_y \\ \mathbf{H}_z \end{bmatrix} \mathbf{K}_x^2, \tag{30}$$

which defines \mathbf{K}_x^2 and $[\mathbf{H}_y, \mathbf{H}_z]^T$ as a solution of the second-order generalized eigenvalue problem (GEP).

2.3 Obtaining electrical fields

For each eigenvalue there exist two propagation constants $\pm \mathbf{K}_x$. In order to have physically meaningful propagation constants, the imaginary part of the elements of \mathbf{K}_x are chosen to be positive, which corresponds to a forward propagating wave; then $-\mathbf{K}_x$ denotes the backward propagating wave.

We also write the electrical field components as Bloch modes,

$$E_i = \mathbf{v}^T \mathbf{E}_i \mathbf{v}, \quad i = x, y, z. \tag{31}$$

From Maxwell's curl equations, $\nabla \times \vec{E} = j\omega\mu\vec{H}$ and $\nabla \times \vec{H} = -j\omega\epsilon\vec{E}$, we get

$$\partial_x E_y = j\omega\mu H_z - \partial_y [(j\omega\epsilon)^{-1} (\partial_y H_z - \partial_z H_y)], \tag{32}$$

$$\partial_x E_z = j\omega\mu H_y - \partial_z [(j\omega\epsilon)^{-1} (\partial_y H_z - \partial_z H_y)]. \tag{33}$$

Using the relations in (8), (31), we can rewrite (32) and (33) as

$$\eta_0^{-1} \mathbf{v}^T \mathbf{K}_x \mathbf{E}_y \mathbf{v} = \mathbf{v}^T [\mathbf{K}_y \tilde{\Gamma} \mathbf{K}_z \mathbf{H}_y + (\mathbf{I} - \mathbf{K}_y \tilde{\Gamma} \mathbf{K}_y) \mathbf{H}_z] \mathbf{v}, \tag{34}$$

$$\eta_0^{-1} \mathbf{v}^T \mathbf{K}_x \mathbf{E}_z \mathbf{v} = \mathbf{v}^T [(-\mathbf{I} + \mathbf{K}_z \tilde{\Gamma} \mathbf{K}_z) \mathbf{H}_y - \mathbf{K}_z \tilde{\Gamma} \mathbf{K}_y \mathbf{H}_z] \mathbf{v}. \tag{35}$$

The coefficients of each mode should be matched, thus scalar equations, (34) and (35), can be rewritten as the following matrix equation:

$$\begin{bmatrix} \mathbf{E}_y \\ \mathbf{E}_z \end{bmatrix} = \eta_0 \begin{bmatrix} \mathbf{K}_y \tilde{\Gamma} \mathbf{K}_z & \mathbf{I} - \mathbf{K}_y \tilde{\Gamma} \mathbf{K}_y \\ -\mathbf{I} + \mathbf{K}_z \tilde{\Gamma} \mathbf{K}_z & \mathbf{K}_z \tilde{\Gamma} \mathbf{K}_y \end{bmatrix} \times \mathbf{K}_x^{-1} \begin{bmatrix} \mathbf{H}_y \\ \mathbf{H}_z \end{bmatrix}. \tag{36}$$

2.4 Fast-convergence formulation

Many researchers [19, 30, 31] have found that there is a problem in the Fourier expansion due to discontinuities in the material dielectric function which causes a slow convergence in the algorithm. Li introduced an inverse rule for the Fourier expansion of the product of two functions $f(x)$ and $g(x)$ [20]

$$[\mathcal{F}_N[f(x)g(x)]]_n = \sum_{m=-M}^M \left[\mathcal{F}_L \left[\frac{1}{f(x)} \right] \right]_{nm}^{-1} [\mathcal{F}_M[g(x)]]_m, \tag{37}$$

where $\mathcal{F}_N[\cdot]$ denotes the truncated Fourier coefficients of a function, e.g. $f(x) = \sum_{n=-N}^N [\mathcal{F}_N[f(x)]]_n \exp(jnx)$, and $[\mathbf{v}]_{pq}$ denotes the (p, q) th entry of the matrix generated from the elements of vector \mathbf{v} . Its value equals the $(p - q)$ th element of the vector whose index ranges from $-\text{Max}(q)$ to $\text{Max}(p)$. It is stated in [20] that two piecewise-smooth, bounded, periodic functions that have only pairwise-complementary jump discontinuities should be Fourier expanded by the inverse rule rather than Laurent’s rule.

To apply the new Fourier expansion law in our algorithm, we need to introduce the following new matrices. We apply

$$[\tilde{\Gamma}_y(z)]_{mn} = \frac{1}{l_y} \int_0^{l_y} \frac{1}{\epsilon_r} \exp[-j(m - n)\bar{l}_y y] dy, \tag{38}$$

$$[\tilde{\Gamma}_z(y)]_{pq} = \frac{1}{l_z} \int_0^{l_z} \frac{1}{\epsilon_r} \exp[-j(m - n)\bar{l}_z z] dz, \tag{39}$$

where $\bar{l}_y = 2\pi/l_y$ and $\bar{l}_z = 2\pi/l_z$. Obviously Γ_y and Γ_z are still functions of z and y , respectively

$$[\Gamma_{yz}]_{rs} = \frac{1}{l_z} \int_0^{l_z} [[\tilde{\Gamma}_y(z)]^{-1}]_{mp} \exp[-j(n - q)\bar{l}_z z] dz, \tag{40}$$

$$[\Gamma_{zy}]_{rs} = \frac{1}{l_y} \int_0^{l_y} [[\tilde{\Gamma}_z(y)]^{-1}]_{nq} \exp[-j(m - p)\bar{l}_y y] dy, \tag{41}$$

where the relation between r, s and m, n, p, q are defined in (24) and (25). Then, we should modify (30) to be

$$\begin{bmatrix} \Gamma_{zy} - \mathbf{K}_y^2 - \Gamma_{zy} \mathbf{K}_z \Gamma^{-1} \mathbf{K}_z & \Gamma_{zy} \mathbf{K}_z \Gamma^{-1} \mathbf{K}_y - \mathbf{K}_y \mathbf{K}_z \\ \Gamma_{yz} \mathbf{K}_y \Gamma^{-1} \mathbf{K}_z - \mathbf{K}_z \mathbf{K}_y & \Gamma_{yz} - \mathbf{K}_z^2 - \Gamma_{yz} \mathbf{K}_y \Gamma^{-1} \mathbf{K}_y \end{bmatrix} \times \begin{bmatrix} \mathbf{H}_y \\ \mathbf{H}_z \end{bmatrix} = \begin{bmatrix} \mathbf{H}_y \\ \mathbf{H}_z \end{bmatrix} \mathbf{K}_x^2, \tag{42}$$

which defines \mathbf{K}_x^2 and $[\mathbf{H}_y, \mathbf{H}_z]^T$ as a solution of the second-order standard eigenvalue problem (SEP). Similarly, we should also change the electrical field expression in (36) to

$$\begin{bmatrix} \mathbf{E}_y \\ \mathbf{E}_z \end{bmatrix} = \eta_0 \begin{bmatrix} \mathbf{K}_y \Gamma^{-1} \mathbf{K}_z & \mathbf{I} - \mathbf{K}_y \Gamma^{-1} \mathbf{K}_y \\ -\mathbf{I} + \mathbf{K}_z \Gamma^{-1} \mathbf{K}_z & \mathbf{K}_z \Gamma^{-1} \mathbf{K}_y \end{bmatrix} \times \mathbf{K}_x^{-1} \begin{bmatrix} \mathbf{H}_y \\ \mathbf{H}_z \end{bmatrix}. \tag{43}$$

Other methods for improving the convergence have been studied [32–39]. Granet and Plumey showed that an appropriately designed non-linear coordinate can increase the spatial resolution around discontinuities, therefore reducing the number of modes needed to achieve convergence [35]. Popov and Neviere proposed to use the local tangential and normal fields in calculations to improve convergence [32–34, 36], and their method has been shown to be effective by various studies [37–39]. However, these methods all involve coordinate transformations, which are structure-dependent, hence non-universal and non-unique. Furthermore, if a multilayer structure contains different geometries in different layers, then different transformations are required, which makes cascading very difficult. Thus these methods are not discussed in this paper.

2.5 Cascading layers using S-matrix algorithm

Note that the scattering matrix algorithm we presented in this section is mainly based on the papers [24, 25]. Suppose we have an N -layer structure as shown in Fig. 3. The layers are indicated as Region 1 to Region N in the figure; Region 0 and Region $N + 1$ is the substrate and the superstrate, respectively. We introduce a notation for the expansion of the

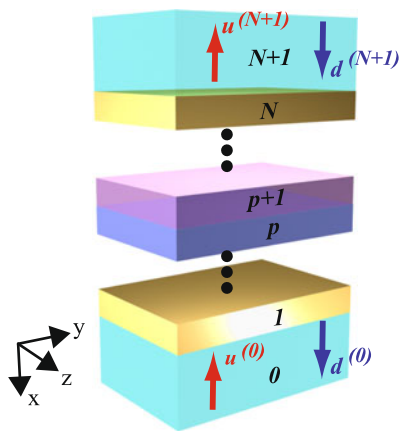


Fig. 3 Abstract view of a multilayer structure with substrate in region 0 and superstrate in region $N + 1$

transverse fields

$$\mathbf{H}_t^{(p)} = \begin{bmatrix} \mathbf{H}_y^{(p)} \\ \mathbf{H}_z^{(p)} \end{bmatrix}, \quad \mathbf{E}_t^{(p)} = \begin{bmatrix} \mathbf{E}_y^{(p)} \\ \mathbf{E}_z^{(p)} \end{bmatrix}, \quad (44)$$

where the superscript (p) denotes the value for the p th layer. δ is the thickness of the layer. Applying the boundary conditions on the interface between Layer p and Layer $p + 1$, we have

$$\begin{bmatrix} \mathbf{E}_t^{(p+1)} & \mathbf{E}_t^{(p+1)} \\ \mathbf{H}_t^{(p+1)} & -\mathbf{H}_t^{(p+1)} \end{bmatrix} \begin{bmatrix} \mathbf{u}^{(p+1)} \\ \mathbf{d}^{(p+1)} \end{bmatrix} = \begin{bmatrix} \mathbf{E}_t^{(p)} & \mathbf{E}_t^{(p)} \\ \mathbf{H}_t^{(p)} & -\mathbf{H}_t^{(p)} \end{bmatrix} \times \begin{bmatrix} \exp(j\mathbf{K}_x^{(p)}\delta^{(p)}) & \mathbf{O} \\ \mathbf{O} & \exp(-j\mathbf{K}_x^{(p)}\delta^{(p)}) \end{bmatrix} \begin{bmatrix} \mathbf{u}^{(p)} \\ \mathbf{d}^{(p)} \end{bmatrix}, \quad (45)$$

where vectors \mathbf{u} and \mathbf{d} are the coefficients of the eigenmodes of upward wave and downward wave, respectively. The exponential operator for a diagonal matrix means exponentiating every entry on the main diagonal. For any $0 \leq p \leq N$, matrix $\mathbf{S}^{(p)}$ links the waves in layer $p + 1$ and medium 0 in

$$\begin{bmatrix} \mathbf{u}^{(p+1)} \\ \mathbf{d}^{(0)} \end{bmatrix} = \mathbf{S}^{(p)} \begin{bmatrix} \mathbf{u}^{(0)} \\ \mathbf{d}^{(p+1)} \end{bmatrix}, \quad (46)$$

where the matrix $\mathbf{S}^{(p)}$ can be written as

$$\mathbf{S}^{(p)} = \begin{bmatrix} \mathbf{T}_{uu}^{(p)} & \mathbf{R}_{ud}^{(p)} \\ \mathbf{R}_{du}^{(p)} & \mathbf{T}_{dd}^{(p)} \end{bmatrix}, \quad (47)$$

and the four sub-matrices have the following physical meanings; \mathbf{T} is the transmission matrix, \mathbf{R} is the reflection matrix, and the subscripts denote which waves those matrices

link (u denotes an upward wave and d denotes a downward wave).

Combining (45) and (47) to eliminate $\mathbf{u}^{(p)}$ and $\mathbf{d}^{(p)}$, we obtain recursive expressions for the sub-matrices of $\mathbf{S}^{(p)}$

$$\mathbf{T}_{uu}^{(p)} = [\mathbf{t}_1^{(p)} - \mathbf{R}_{ud}^{(p)}\mathbf{t}_2^{(p)}]\Phi_x^{(p)}\mathbf{T}_{uu}^{(p-1)}, \quad (48)$$

$$\mathbf{R}_{ud}^{(p)} = [\mathbf{t}_2^{(p)} + \mathbf{t}_1\Omega^{(p)}][\mathbf{t}_1^{(p)} + \mathbf{t}_2\Omega^{(p)}]^{-1}, \quad (49)$$

$$\mathbf{R}_{du}^{(p)} = \mathbf{R}_{du}^{(p-1)} - \mathbf{T}_{dd}^{(p)}\mathbf{t}_2^{(p)}\Phi_x^{(p)}\mathbf{T}_{uu}^{(p-1)}, \quad (50)$$

$$\mathbf{T}_{dd}^{(p)} = \mathbf{T}_{dd}^{(p-1)}\Phi_x^{(p)}[\mathbf{t}_1^{(p)} + \mathbf{t}_2\Omega^{(p)}]^{-1}, \quad (51)$$

where

$$\mathbf{t}_1^{(p)} = \frac{1}{2}[(\mathbf{E}_t^{(p+1)})^{-1}\mathbf{E}_t^{(p)} + (\mathbf{H}_t^{(p+1)})^{-1}\mathbf{H}_t^{(p)}], \quad (52)$$

$$\mathbf{t}_2^{(p)} = \frac{1}{2}[(\mathbf{E}_t^{(p+1)})^{-1}\mathbf{E}_t^{(p)} - (\mathbf{H}_t^{(p+1)})^{-1}\mathbf{H}_t^{(p)}], \quad (53)$$

$$\Omega^{(p)} = \Phi_x^{(p)}\mathbf{R}_{ud}^{(p-1)}\Phi_x^{(p)}, \quad (54)$$

$$\Phi_x^{(p)} = \exp(j\mathbf{K}_x^{(p)}\delta^{(p)}). \quad (55)$$

To start the recursion, we can set

$$\mathbf{S}^{(-1)} = \begin{bmatrix} \mathbf{I} & \mathbf{O} \\ \mathbf{O} & \mathbf{I} \end{bmatrix}, \quad (56)$$

which means if we add a virtual layer between Layer 0 and Layer 1, and the thickness of the layer is zero, then there is no reflection on the interface. So the reflection matrices are \mathbf{O} and transmission matrices are \mathbf{I} . Since Layer 0 is homogeneous, $\mathbf{H}_t^{(0)} = \mathbf{I}$ and

$$[\mathbf{K}_x^{(0)}]^2 = n_{\text{sub}}^2\mathbf{I} - \mathbf{K}_y^2 - \mathbf{K}_z^2, \quad (57)$$

where n_{sub} is the reflective index of the substrate. Similarly, in the Layer $N + 1$, we have $\mathbf{H}_t^{(N+1)} = \mathbf{I}$ and

$$[\mathbf{K}_x^{(N+1)}]^2 = n_{\text{sup}}^2\mathbf{I} - \mathbf{K}_y^2 - \mathbf{K}_z^2, \quad (58)$$

where n_{sup} is the reflective index of the superstrate.

2.6 Diffraction coefficients for plane wave incidence

For plane-wave incidence as described in Sect. 2.1 on an N -layered structure as shown in Fig. 3, we can easily find the following conditions:

$$\mathbf{u}^{(0)} = [\underbrace{0, 0, \dots, 0}_{2(2M_y+1)(2M_z+1)}]^T, \quad (59)$$

$$\mathbf{d}^{(N+1)} = [\underbrace{0, \dots, 0}_{2M_y M_z + M_y + M_z}, u_y, \underbrace{0, \dots, 0}_{2M_y M_z + M_y + M_z}, \underbrace{0, \dots, 0}_{2M_y M_z + M_y + M_z}, u_z, \underbrace{0, \dots, 0}_{2M_y M_z + M_y + M_z}]^T. \quad (60)$$

From (46) with $p = N$ we can get the reflected wave coefficient $\mathbf{u}^{(N+1)}$ and the transmitted wave coefficient $\mathbf{d}^{(0)}$. Without losing generality, we assume a unity incident wave magnitude, thus, $\mathbf{u}^{(N+1)}$ and $\mathbf{d}^{(0)}$ contains all the diffraction coefficients for all diffraction orders. Especially, the $(2M_y M_z + M_y + M_z + 1)$ th element and $(6M_y M_z + 2M_y + 2M_z + 2)$ th element of $\mathbf{u}^{(N+1)}$ are the zero-order reflection coefficients for y and z components, respectively. Similarly, the elements in the same positions of $\mathbf{d}^{(0)}$ are the zero-order transmission coefficients for the y and z components.

3 Numerical tests and discussions

3.1 Implementation details

Both the initial SHA formulation and the fast-convergence formulation were implemented in C language in our tests. The solver was compiled using Intel C Compiler version 11.0.083 and linked with Intel Math Kernel Library (MKL) version 10.0.3.020 which contains both optimized Lapack and ScaLapack for Intel platforms. It was parallelized using Message Passing Interface (MPI) version 2 and ScaLapack for out-of-core computations. Parallelism in SHA can be achieved through distributing the independent GEP/SEP problems for multiple layers among available processors (quasi-parallelism) and by scaling the GEP/SEP solver itself. We note that the scalability of the GEP/SEP part itself is poor due to a known issue in the complex eigenvector calculation subroutine (PZTREV) in ScaLapack. The subroutine is parallelized, but also includes significant sequential tasks, preventing efficient scaling. The performance does not scale well on Xeons and Itaniums (with any OS), as shown in Fig. 4. The picture is slightly better for the Intel Core i7 processors series. Thus, although the parallel performance of our solver scales appropriately for cascaded structures with a large number of layers (approaching a linear speedup), further increasing the efficiency and scalability of GEP/SEP requires substantial work going beyond the scope of this study.

The compiling and testing environment was dual-CPU Quad-Core Intel Xeon E5410 2.33 GHz cluster nodes with 16 GB memory and Gigabit Ethernet interconnects. The typical computation time per wavelength and memory consumption for different numbers of modes is shown in Table 1. The numbers, such as 25×25 , indicate the number of modes used for y and z directions, namely M_y and M_z in the formulation, respectively. We note that the actual system matrix dimension in the eigenvalue problem is $2(2M_y + 1)(2M_z + 1)$ and the matrix is dense.

We note that a lightweight (but fully-functional) version of an SHA solver for two-dimensional multilayer structures (1-D metal-dielectric gratings) with a user-friendly interface

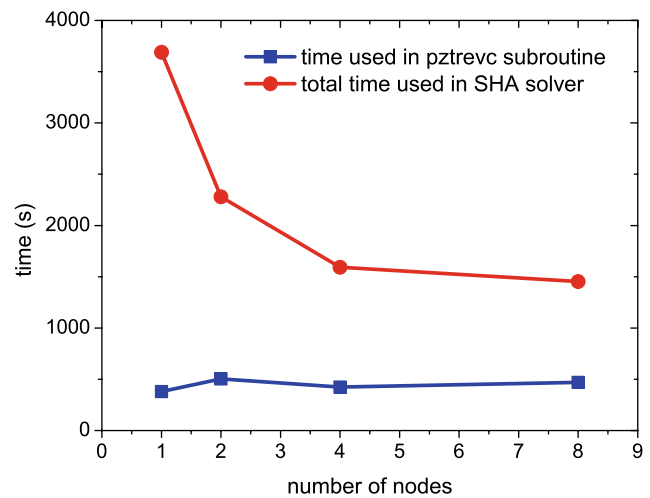


Fig. 4 Time number of processors using GEP subroutines in ScaLapack. 25×25 modes (matrix dimension is 5202) was used in this test

Table 1 The typical time and memory consumption for different number of modes

| Modes | Matrix dimension | Memory | Time |
|----------------|------------------|--------|--------|
| 20×20 | 3362 | 635 MB | 6 min |
| 25×25 | 5202 | 1.4 GB | 20 min |
| 30×30 | 7442 | 2.8 GB | 1.0 h |
| 35×35 | 10082 | 5.0 GB | 2.39 h |
| 40×40 | 13122 | 8.4 GB | 5.2 h |

has been staged online by our group and is free to use [40]. The solver uses optical constants linked to [41] and therefore can utilize tabulated data for a wide range of dispersive materials, including metals.

3.2 Validation test

We used the simulation method on a gold nanoantenna array designed in [42] as a validation test. A unit cell of the array is composed of gold nanoparticles with elliptic cylinder shapes, as illustrated in Fig. 5(a), and a gap between the two particles along their major axes. Silica is used as the substrate. The structure was fabricated and examined by field emission scanning electron microscopy (FESEM), and a representative FESEM image is shown in Fig. 5(b). Although all unit cells in an array have the same design, each particle randomly deviates from the designed shape to certain extent. In the sample shown, the major axis of the nanoparticles ranges from 104 to 118 nm, the minor axis from 52 to 62 nm, and the gap from 12 to 27 nm. The transmittance and reflectance spectra of the sample was measured in the visible range with incident light normal to the sample surface. The incident light, which was illuminated along x direction, was linearly polarized as either the principal polarization, where the electric field is parallel to the major

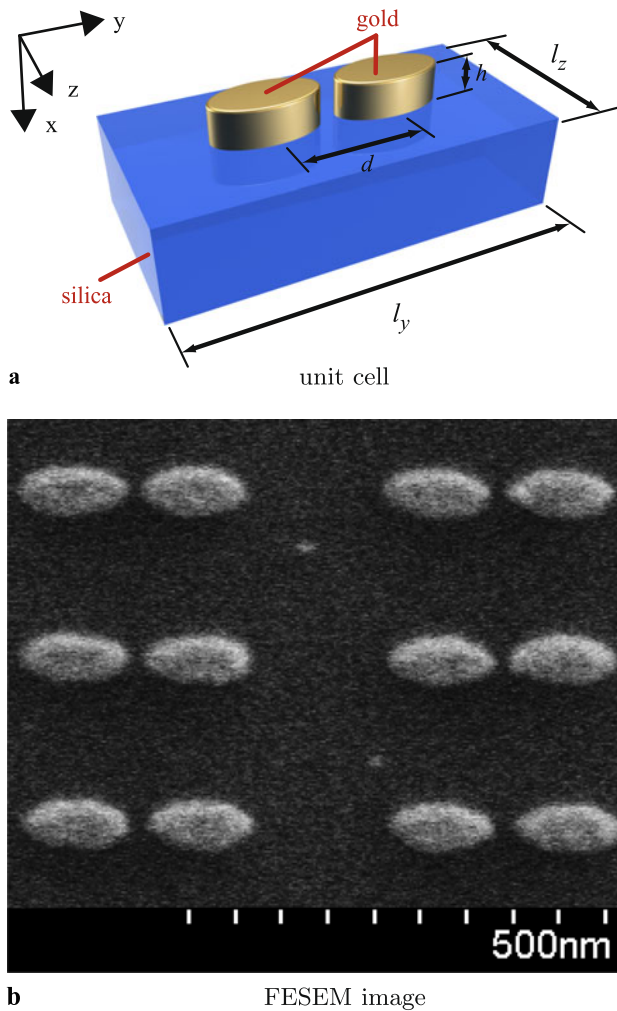


Fig. 5 (a) A unit cell of the gold nanoantenna array on the silica substrate; (b) FESEM image of the sample of the gold nanoantenna array

axis, or the secondary polarization, where the electric field is normal to the major axis. The structural dimensions used in simulations are as follows: the major axis of the gold cylinder is 110 nm, the minor axis is 55 nm, the gap is 17 nm, the antenna thickness is 40 nm. These values are all within the range of dimensions obtained from FESEM measurements. The thickness of the antennas, $h = 40$ nm, the period along the major axis, $l_y = 400$ nm, and along the minor axis, $l_z = 200$ nm, were taken from the initial design. The small size of the metallic nanoparticles and a large electric resonance due to the small gap between two nanoparticles are among the main challenges of the test structure.

To obtain a good set of reference data, we also simulated the structure using a commercial finite-element method (FEM) software package, COMSOL Multiphysics, with 3-order elements. The validity of the FEM solution was verified by using the same model with different levels of additional meshing refinement and an adaptive solver.

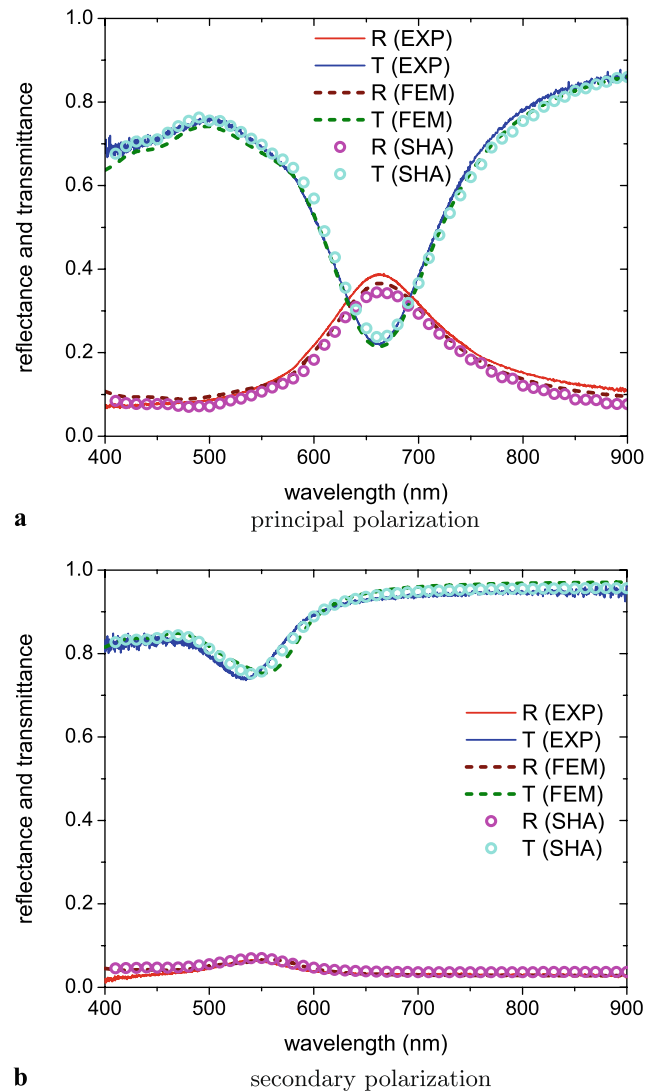


Fig. 6 Comparison of the reflection and transmission spectra obtained in experiments and simulations using FEM and SHA in two polarizations: (a) principal polarization and (b) secondary polarization

A best-fit Drude–Lorentz model for experimental data of gold from Johnson and Christy [43] was used in the simulation. A loss factor of 3 was applied to the Drude–Lorentz model in order to take size effect and the internal grain structure in the fabricated sample into account [42, 44]. The Drude–Lorentz model parameters used in this paper can be found in the online tool, PhotonicsDB, developed by our group [41]. The simulation results are compared with the experimental spectra in Fig. 6. The test structure of Fig. 5 reveals all typical features of localized plasmonic resonances, e.g. the spectra show a strong resonance for the principal polarization at 660 nm. We can see that our approach converges well in both principal polarization and secondary polarization spectra, and the difference among experimental results, FEM results and SHA results is almost indiscernible.

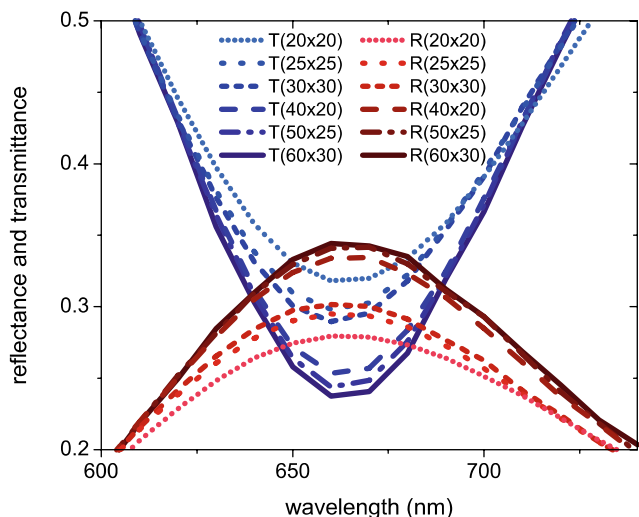


Fig. 7 Reflection and transmission spectra near the resonance wavelength calculated with a different number of spatial harmonics

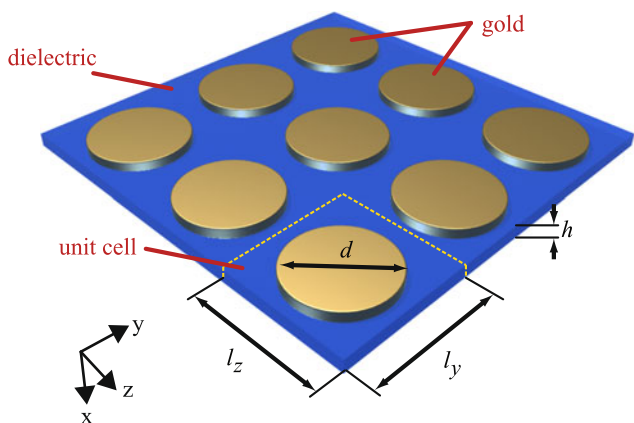
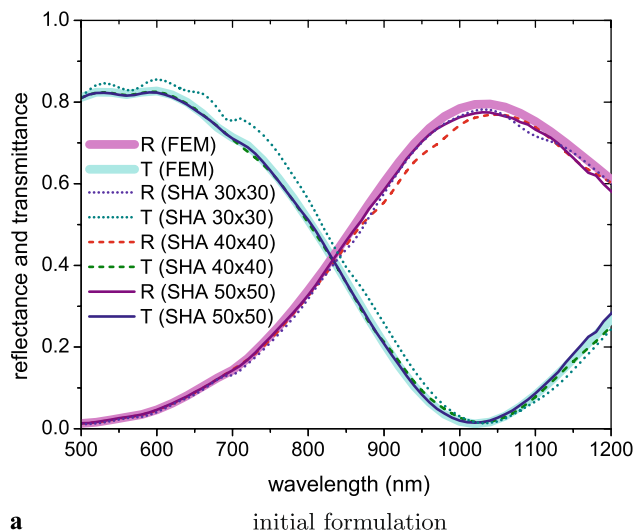


Fig. 8 Schematic view of gold nano-disk array structure

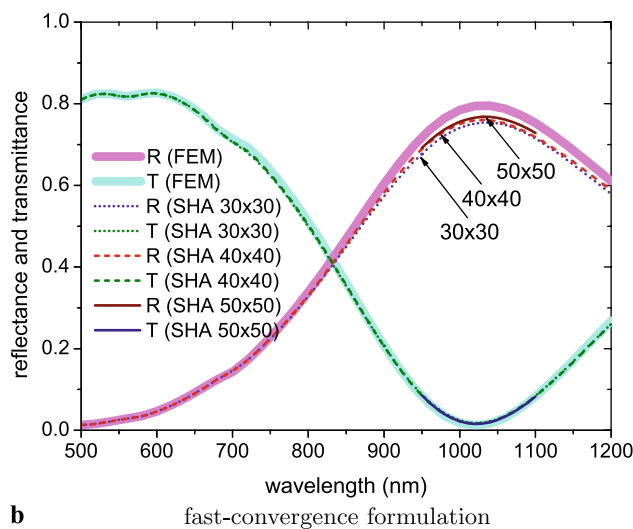
3.3 Convergence test

In our test, we observe that the off-resonance portion of the spectra will converge even using a relatively small number of spatial harmonics. However, the spectra near the resonance wavelength require a large number of modes in calculation to converge due to the strong near-field coupling between the nanoparticles. Figure 7 shows the reflection and transmission spectra near the resonance wavelength using the fast-convergence formulation described in Sect. 2.4 with different numbers of spatial harmonics. We can see both reflection and transmission spectra converge uniformly.

In this particular example, there is not much difference in convergence between a direct application of the initial formulation (30) and the fast-convergence formulation (42). In order to make the difference more clear, we used another metal-dielectric nanostructure which is shown in Fig. 8. This structure has only one layer and has semi-infinite air regions at both ends. Cylindrical gold nano-disks are arranged



a



b

Fig. 9 Comparison of the speed of convergence of reflection and transmission spectra obtained in SHA simulations using (a) initial formulation and (b) fast-convergence formulation

in a two-dimensional array which is embedded in a dielectric host with refractive index $n = 1.45$. The period in the y direction is $l_y = 400$ nm, and the period in the z direction is $l_z = 400$ nm. The diameter of each gold nano-disk is $d = 300$ nm, and the thickness of the layer is 10 nm. The incident light, which is illuminated along the x direction, was linearly polarized in either the y or z direction. The structure is intentionally designed to be symmetric in the y and z directions in order to be able to use the same modes in both y and z directions. The large filling factor of the metallic nano-disks and a broad electric resonance at a wavelength of about 1 micron are among the most difficult aspects of this model.

Figure 9 shows the convergence of the solutions for the gold nano-disk array structure while increasing the number of modes simultaneously in both the y and z directions using

the initial formulation (Fig. 9(a)) and fast-convergence formulation (Fig. 9(b)). The FEM results are also included as a reference. From Fig. 9(a) we can obviously see that when we have 30×30 modes, the spectra matches the FEM result better than with 40×40 modes, in the long wavelength region, while it has a much worse match in the short wavelength region. In addition, we can also clearly see that there is a small feature in the spectra of 30×30 modes at around 1100 nm which is an artifact. when we increase the number of modes, the curve becomes smooth and has a good match with the FEM results. In contrast, if we calculate using fast-convergence formulation, all spectra are smooth even when using less modes, and they approach the FEM reference result uniformly as we increase modes, as shown in Fig. 9(b). In summary, the fast-convergence formulation shows a more uniform convergence. A direct application of the initial formulation is of limited utility to the problem due to the fact that the convergence is not uniform, and that there may be artifacts which need a substantial number of spatial modes to resolve.

4 Conclusion

A second-order formulation of the SHA method for three-dimensional structures (bi-periodic multilayer structures) has been presented. The scattering matrix algorithm is used to cascade multiple layers. We have calculated the transmittance and reflectance spectra of a nanoantenna sample using SHA, and the results match those obtained from experiments and from finite-element method simulations very well. A study of the convergence of the initial formulation and fast-convergence formulation shows that the latter formulation has a more uniform convergence while handling metal-dielectric structures.

Acknowledgements We would like to cite fruitful discussions of SHA fundamentals with U. Chettiar. Comments regarding scalable implementation (A. Sameh and Intel MKL Support Team) are greatly appreciated. FESEM image of the sample was done by H.-K. Yuan. We are also happy have had kind assistance from G. Klimeck (Network for Computational Nanotechnology), W. Whitson and D. Seaman (Rosen Center for Advanced Computing). Finally, we would like to thank our group leader V. Shalaev for his patience and unconditional support to the development of this important tool.

This work was supported in part by ARO MURI grant 50342-PH-MUR and the Danish Research Council for Technology and Production Sciences (grant 274-07-0057).

References

1. R.E. Collin, *Can. J. Phys.* **34**, 398–411 (1956)
2. C.B. Burckhardt, *J. Opt. Soc. Am.* **56**(11), 1502–1508 (1966)
3. H. Kogelnik, *Bell Syst. Tech. J.* **48**(9), 2909–2947 (1969)
4. F.G. Kaspar, *J. Opt. Soc. Am.* **63**(1), 37–45 (1973)
5. R. Magnusson, T.K. Gaylord, *J. Opt. Soc. Am.* **67**(9), 1165–1170 (1977)
6. K. Knop, *J. Opt. Soc. Am.* **68**(9), 1206–1210 (1978)
7. I.C. Botten, M.S. Craig, R.C. McPhedran, J.L. Adams, J.R. Andrewartha, *J. Mod. Opt.* **28**(3), 413–428 (1981)
8. L.C. Botten, M.S. Craig, *Opt. Acta* **28**(8), 1103–1106 (1981)
9. L.C. Botten, M.S. Craig, R.C. McPhedran, J.L. Adams, J.R. Andrewartha, *J. Mod. Opt.* **28**, 1087–1102 (1981)
10. M.G. Moharam, T.K. Gaylord, *J. Opt. Soc. Am.* **71**(7), 811–818 (1981)
11. M.G. Moharam, T.K. Gaylord, *J. Opt. Soc. Am.* **73**(9), 1105–1112 (1983)
12. M.G. Moharam, T.K. Gaylord, *J. Opt. Soc. Am.* **73**, 1105–1112 (1983) (1917–1983)
13. L. Li, *J. Opt. Soc. Am. A* **10**(12), 2581–2591 (1993)
14. E. Noponen, J. Turunen, *J. Opt. Soc. Am. A* **11**(9), 2494–2502 (1994)
15. P. Lalanne, *J. Opt. Soc. Am. A* **14**(7), 1592–1598 (1997)
16. B. Momeni, B. Rashidian, *IEEE Trans. Antennas Propag.* **52**(1), 165–171 (2004)
17. P. Lalanne, G.M. Morris, *J. Opt. Soc. Am. A* **13**(4), 779–784 (1996)
18. G. Granet, B. Guizal, *J. Opt. Soc. Am. A* **13**(5), 1019–1023 (1996)
19. J.M. Jarem, P.P. Banerjee, *Computational Methods for Electromagnetic and Optical Systems* (Dekker, New York, 2000)
20. L. Li, *J. Opt. Soc. Am. A-Opt. Image Sci. Vis.* **13**(9), 1870–1876 (1996)
21. L. Li, *J. Opt. Soc. Am. A* **10**(12), 2581–2591 (1993)
22. F. Montiel, M. Nevière, *J. Opt. Soc. Am. A* **11**(12), 3241–3250 (1994)
23. N.P.K. Cotter, T.W. Preist, J.R. Sambles, *J. Opt. Soc. Am. A* **12**(5), 1097–1103 (1995)
24. L. Li, *J. Opt. Soc. Am. A* **13**(5), 1024–1035 (1996)
25. L. Li, *J. Opt. Soc. Am. A* **20**(4), 655–660 (2003)
26. A.V. Kildishev, U.K. Chettiar, *J. Appl. Comput. Electromagn. Soc.* **22**, 172–183 (2007)
27. C. Helgert, C. Menzel, C. Rockstuhl, E. Pshenay-Severin, E.-B. Kley, A. Chipouline, A. Tünnermann, F. Lederer, T. Pertsch, *Opt. Lett.* **34**(5), 704–706 (2009)
28. D.W. Peters, D.J. Shelton, J.C. Ginn, M.B. Sinclair, G.A. TenEyck, J. Wendt, W.L. Langston, L. Basilio, L. Warne, B.S. Passmore, I. Brener, G.D. Boreman, in *3rd International Congress on Advanced Electromagnetic Materials in Microwaves and Optics*, London, UK, 2009
29. Z. Ku, J. Zhang, S.R.J. Brueck, *Opt. Express* **17**(8), 6782–6789 (2009)
30. L. Li, J. Chandezon, *J. Opt. Soc. Am. A-Opt. Image Sci. Vis.* **13**(11), 2247–2255 (1996)
31. L. Li, *J. Opt. Soc. Am. A-Opt. Image Sci. Vis.* **14**(10), 2758–2767 (1997)
32. E. Popov, M. Nevière, *Opt. Lett.* **25**(9), 598–600 (2000)
33. E. Popov, M. Nevière, *J. Opt. Soc. Am. A* **17**(10), 1773–1784 (2000)
34. E. Popov, M. Nevière, *J. Opt. Soc. Am. A* **18**(11), 2886–2894 (2001)
35. G. Granet, J.-P. Plumey, *J. Opt. A, Pure Appl. Opt.* **4**(5), S145–S149 (2002)
36. E. Popov, M. Nevière, B. Gralak, G. Tayeb, *J. Opt. Soc. Am. A* **19**(1), 33–42 (2002)
37. A. David, H. Benisty, C. Weisbuch, *Phys. Rev. B* **73**(7), 075107 (2006)
38. T. Schuster, J. Ruoff, N. Kerwien, S. Rafler, W. Osten, *J. Opt. Soc. Am. A* **24**(9), 2880–2890 (2007)
39. R. Antos, *Opt. Express* **17**(9), 7269–7274 (2009)
40. X. Ni, Z. Liu, F. Gu, M.G. Pacheco, A.V. Kildishev, doi:[10254/nanohub-r6977.6](https://doi.org/10.254/nanohub-r6977.6) (2009)

41. X. Ni, Z. Liu, A.V. Kildishev, doi:[10254/nanohub-r3692.6](https://doi.org/10.254/nanohub-r3692.6) (2008)
42. Z. Liu, A. Boltasseva, R.H. Pedersen, R. Bakker, A.V. Kildishev, V.P. Drachev, V.M. Shalaev, *Metamaterials* **2**, 45–51 (2008)
43. P.B. Johnson, R.W. Christy, *Phys. Rev. B* **6**, 4370–4379 (1972)
44. J. Borneman, A.V. Kildishev, K.-P. Chen, V.P. Drachev, in *COM-SOL Conference*, Boston, USA, 2009