Model Order Reduction of Large-Scale Metasurfaces Using a Hierarchical Dipole Approximation

Mohammad Mahdi Salary, Ali Forouzmand, and Hossein Mosallaei*

Department of Electrical and Computer Engineering, Northeastern University, Boston, Massachusetts 02115, United States

Supporting Information

ABSTRACT: Advances in the field of metasurfaces require simulation of large-scale metasurfaces that extend over many light wavelengths. Adopting standard numerical methods leads to models featuring a large number of degrees of freedom, which are prohibitive to solve within a time window compatible with the design workflow. Therefore, this demands developing the techniques to replace large-scale computational models with simpler ones, still capable of capturing the essential features but imposing a fraction of the initial computational costs. In this work, we present a simulation approach in order to handle reduced order analyses of large-scale metasurfaces of arbitrary elements. We use the discrete dipole approximation in conjunction with the discrete complex image method and hierarchical matrix construction as a common theoretical framework for dipole approximation in the hierarchy of individual elements and the array scale. We extract the contributions of multipoles in the scattering spectra of the nanoantennas forming the metasurface and retrieve their dynamic polarizabilities. The computational complexity of modeling the array problem is then significantly reduced by replacing the fine meshing of each nanoantenna with its dynamic polarizability. The solver is developed to model several fully functional metasurfaces of different types including a one-atom-thick metasurface made of graphene with chemical doping interruptions, a multifocusing lens made of plasmonic V-shaped nanoantennas, and a multicolor hologram consisting of dielectric nanobars. The performance of the method is evaluated through comparison with full-wave simulations, and a significant computational gain is observed while the accuracy of the results is retained owing to the preserved coupling information between dipolar modes.

KEYWORDS: metasurfaces, polarizability retrieval, discrete dipole approximation, nanoantennas, meta-lens, meta-hologram

Over the past decade, metamaterials have emerged as a rapidly growing interdisciplinary area involving physics, chemistry, optics, and nanoscience. Metamaterials, composed of artificial nanostructures, can be engineered to possess arbitrary effective permittivity and permeability unattainable in the conventional materials and thus offer a great platform for extending functionality of devices and achieving ultimate control of light in the nanoscale dimension. A great deal of attention has been paid to ultrathin flat metamaterials, or metasurfaces, because of their low profile and ease of realization with respect to 3D bulk structures. The metasurfaces can provide a great flexibility in controlling the behavior of reflected and transmitted light by using spatially abrupt changes in the phase response of each building block. Moreover, graded index metasurfaces offer a promising solution for photonic-integrated circuits through manipulation of surface plasmons.

Several research efforts have been put into designing and building unique metasurfaces at mid-IR, telecommunication, and visible wavelengths, leading to many interesting applications, such as beam steering, focusing lenses, flat-top generators, quarter waveplates, optical vortex plates, and holograms.

From the designer’s perspective, a comprehensive understanding of the scattering properties in large-scale arrays is crucial to build metasurfaces with the desired functionality. As such, there is a need for accurate and efficient methods of simulations for a given structure. For design purposes, purely numerical EM solvers such as the finite element method (FEM), the finite-difference time-domain (FDTD) method, the...
method of moments (MoM), and the discrete-dipole approximation (DDA) are employed in the periodic arrangement to obtain the local periodicity prescription. The most common technique being used is FDTD, which is favored in treating the wide-band response of periodic arrangements. However, functional metasurfaces are inherently aperiodic, as they need to realize a spatial gradient of phase discontinuity or refractive index profile, and the realistic design of metasurfaces requires a fast feedback loop between the local periodicity prescription and the aperiodic results, allowing an iterative process to optimize the metasurface geometry and layout toward a performance metric. In principle, purely numerical methods can also be used for this purpose; however, huge amounts of memory and computing time are required, which make it prohibitive. The state-of-the-art in simulation of aperiodic arrays is the surface integral equation formulation of MoM incorporating modern acceleration techniques such as fast multipole-based methods, fast-Fourier-transform-based methods, and fast low-rank compression methods, which have been recently applied for modeling large arrays with graded patterns.

The alternative to the standard numerical methods is modal analysis techniques including dipolar equations, generalized multipole methods, and T-matrix. These techniques are superior to the other computational methods when the number of elements is large but the number of excited modes in each element could be kept small.

In this context, the dipole modes are the most important of the multipolar modes. All higher order multipole fields decay more strongly with distance than those of the dipoles. In the case of metasurfaces, since the elements are subwavelength, they can typically be well-described as point scatterers with induced dipole moments. The dipole moments describe the response of each element in the array and are proportional to the local field at the phase center of the element with a coupling coefficient, called polarizability. For small scatterers with simple geometries, a closed-form expression for the polarizability may arise from the solution of the Laplace equation in the quasi-static limit. When the inclusions become more complex, dipolar models are still useful, but the dynamic polarizabilities should be determined and numerical efforts are needed. In order to properly take into account collective co- and cross-coupling of electromagnetic fields, the polarizability should be presented in a tensor form.

Recently, the homogenization of metamaterial structures via effective constitutive parameters has received a lot of attention for characterization purposes and as an accurate design means for engineering applications. This objective has led to several techniques for polarizability retrieval of metamaterial inclusions. The equivalent circuit models hold true only in the quasi-static limit and are insufficient to properly take into account retardation or dispersion effects. A number of techniques have also been proposed to convert the results of the quasi-static models into dynamic polarizabilities. This may also fail to predict the correct dynamics of inclusions due to the exotic wave phenomena when the elements are made of plasmonic and high-index dielectric materials. Apart from these techniques, the most popular approach is utilizing transmission and reflection coefficients of the element in a periodic arrangement to retrieve the polarizability. Such approaches may suffer from a variety of artifacts resulting from ignoring the contribution of higher order multipoles. It has been shown that multipolar contributions can influence the polarizability tensor, even in situations where the overall scattering response remains essentially dipolar. As such, the most reliable and rigorous techniques are those based on multipole decomposition and evaluating the dynamic scattering of the element under different measurement conditions.

In this work, we aim at using DDA as a common theoretical framework for dipole approximation in the hierarchy of individual elements and the array scale. The contributions of individual multipole modes to the scattering of each nanomaterial forming the metasurface are extracted, and the polarizability tensors are retrieved by measurements under independent incident conditions. This approach can yield accurate results even for scatterers that poorly verify the quasi-static approximation and take the microscopic structure of arbitrarily shaped nanoantennas into account. The polarizabilities are then used for dipolar analysis in the array scale incorporating theoretical tools developed for acceleration of DDA, namely, the discrete complex image method (DCIM) and hierarchical matrix (H-matrix) construction, which are not met in contemporary multipole solvers such as T-matrix. The proposed simulation approach can be beneficial for modeling different devices in nanophotonics. Specifically, it can be extended to hybrid platforms of metamaterials, quantum dots, and biomolecules due to the exotic wave phenomena when the elements are made of plasmonic and high-index dielectric materials. Apart from these, the most reliable and rigorous techniques are those based on multipole decomposition and evaluating the dynamic scattering of the element under different measurement conditions.

### METHODS

The discrete dipole approximation (also known as the coupled dipole method) is a general computational method based on the volume integral-equation (IE) formulation of Maxwell equations that has been widely used to study scattering problems in electromagnetics and photonics including characterization of the metamaterial elements with complicated shapes. The purpose of this work is to apply the DDA formulation as a common theoretical framework for dipole approximation in the hierarchy of single elements and the array scale.

First, arbitrarily shaped elements are discretized into $N$ cells. In this case, the general formulation of DDA is given by

$$\mathbf{p}_j = \hat{a}_p \mathbf{E}_0(\mathbf{r}_j) + \hat{a}_p \sum_{\mathbf{r}_{\neq j}} G^{\text{EP}}(\mathbf{r}_j, \mathbf{r}_\ell) \mathbf{p}_\ell$$

where $\mathbf{p}_j$ is the induced dipole moment of the $j$th lattice point, $\hat{a}_p$ is the polarizability, $\mathbf{r}_j$ is the position vector of the $j$th lattice point, $\mathbf{E}_0$ is the incident electric field, and $G^{\text{EP}}(\mathbf{r}_j, \mathbf{r}_\ell)$ is the free-space dyadic Green’s function of the surrounding medium without the presence of the scatterers. The polarizability of each cell is expressed by the Clausius–Mossotti relation modified by introducing a nonradiative term, which is known as the lattice dispersion relation (LDR), to improve the convergence.

The DDA can be regarded as valid when the mesh cell size ($d$) is chosen such that $\ln n k_0 d \ll 1$, where $n$ is the complex refractive index of the structure and $k_0$ is the wavenumber in free space.
The dipole moments induced in each lattice point are then used for discrete dipole approximation of individual multipole moments corresponding to the scatterer as

\[ p = \sum_{j=1}^{N} p_j; \hat{\mathbf{Q}}(r_j) = \sum_{j=1}^{N} \hat{\mathbf{Q}}(r_j); \mathbf{m}(r_j) = \sum_{j=1}^{N} m_j(r_j) \]

where

\[ \hat{\mathbf{Q}}(r_j) = \mathbf{3}((r_j - \mathbf{r}_0) \otimes \mathbf{p}) + \mathbf{p} \otimes (r_j - \mathbf{r}_0), \]

\[ m_j(r_j) = -\frac{i\omega}{2}((r_j - \mathbf{r}_0) \times \mathbf{p}) \]

In eqs 2 and 3, \( \omega \) is the angular frequency and \( \otimes \) denotes the dyadic product. As can be seen from the above equation, while the choice of \( \mathbf{r}_0 \) does not affect the electric dipole moment \( \mathbf{p} \), all the higher order multipoles including magnetic moment dipole \( \mathbf{m} \) are dependent upon \( \mathbf{r}_0 \). There exists a point that minimizes the number of required multipoles to describe the scattering behavior. For homogeneous scatterers, this point happens to be the scatterer’s center of mass.\(^{61,62} \) So we define

\[ \mathbf{r}_0 = \frac{1}{N} \sum_{j=1}^{N} r_j/N \]

where \( \mathbf{p}_j \) and \( \mathbf{m}_j \) are induced electric and magnetic dipole moments in the \( j \)th scatterer, and \( \mathbf{E}_0 \) and \( \mathbf{H}_0 \) are the incident electric and magnetic fields.

The periodicity can be considered in the DDA formulism by using the periodic dyadic Green’s functions. For an efficient accurate calculation of the periodic Green’s function the Ewald acceleration method is adopted in this work.\(^{34} \)

The substrate contribution can also be taken into account by incorporating the dyadic Green’s functions of layered media to avoid meshing of the substrate volume. This leads to evaluation of Sommerfeld-type integrals. These integrals are infinite, oscillatory, and slowly convergent integrals, and the numerical integration is very inefficient and time-consuming. In this work, the discrete complex image method (DCIM)\(^ {31,65} \) is applied to reduce the computational cost. The method approximates the integrand by a series of complex exponential functions. The Green’s tensors expressing the contribution of the substrate will then become a series of free-space dyadic Green’s functions with complex distances and amplitudes and can be applied in both periodic and aperiodic cases. Therefore, the computation can be made to be the same order as the free-space problems.

DDA solves a linear system of equations. When a direct method is adopted for this purpose, the computational complexity for a problem with \( N \) unknowns is on the order of \( O(N^3) \) and the memory requirement is proportional to \( O(N^2) \). Employing an iterative solver does not affect the memory requirements. However, the computing time decreases \( O(N_h N^2) \), where \( N_h \) denotes the total number of iterations required to reach convergence. The calculations become quickly prohibitive by increasing \( N \), which hinders applying the method to large-scale problems. Implementing 3D-FFT\(^ {25,60} \)

By choosing six independent incident conditions and using electric and a magnetic dipole response of the element, we are able to accurately retrieve the dynamic polarizability of the element expressed by a \( 6 \times 6 \) polarization tensor:

\[ \mathbf{\hat{\alpha}} = \begin{bmatrix} \hat{\alpha}_E & \hat{\alpha}_{EH} \\ \hat{\alpha}_{HE} & \hat{\alpha}_H \end{bmatrix} = \begin{bmatrix} \mathbf{p}_1 & \ldots & \mathbf{p}_6 \\ \mathbf{m}_1 & \ldots & \mathbf{m}_6 \end{bmatrix} \cdot \begin{bmatrix} \mathbf{E}^{inc}_1 & \ldots & \mathbf{E}^{inc}_6 \\ \mathbf{H}^{inc}_1 & \ldots & \mathbf{H}^{inc}_6 \end{bmatrix}^{-1} \]

This polarizability form takes into account the collective co- and cross-couplings and complex bianisotropic behavior of elements that are influenced by the presence of asymmetry in the geometry of inclusions. \( \hat{\alpha}_E \) and \( \hat{\alpha}_H \) are the usual electric and magnetic polarizability, respectively, whose off-diagonal elements describe the rotational asymmetry of the scatterer. \( \hat{\alpha}_{EH} \) and \( \hat{\alpha}_{HE} \) represent the magnetoelectric response due to mirror asymmetry and directional asymmetry of the element. The retrieved polarizabilities of the elements are then used in an extended DDA formulism known as dual-DDA (D-DDA),\(^ {34} \) which accounts for the contribution of magnetic dipoles and the magnetoelectric cross-couplings:

\[ \mathbf{\mathcal{H}}^{\alpha}_{\mathcal{E}M}(\mathbf{r}_1, \mathbf{r}_2) = \frac{G^{EP}(\mathbf{r}_1, \mathbf{r}_2) G^{EM}(\mathbf{r}_1, \mathbf{r}_2) G^{HP}(\mathbf{r}_1, \mathbf{r}_2) G^{HM}(\mathbf{r}_1, \mathbf{r}_2)}{ \mathbf{m}_1 } \]

and the \( \mathcal{H} \)-matrix\(^ {52} \) can drastically reduce the computation complexity and memory requirements to the order of \( O(N_h N \log(N_h)) \). The \( \mathcal{H} \)-matrix provides a general mathematical framework for low-rank representation of the interaction matrix, consisting of a collection of block matrices of various sizes. This results in compressed storage and reduces the computational complexity of matrix–vector multiplications, which are the bottleneck of iterative solvers. In this work, we have attempted to use the \( \mathcal{H} \)-matrix in dipolar equations for both individual elements and the array scale incorporating the layered media Green’s function. In the high-frequency region (array scale) and in the presence of layered substrates, the low-rank representations cannot be carried out with a constant rank as the substrate contribution, and the domain electrical size significantly affects the rank deficiency. An adaptive rank-revealing algorithm for a prescribed accuracy is required.\(^ {25,26} \)

For this purpose, the adaptive cross approximation (ACA) algorithm\(^ {34} \) is employed to compute low-rank approximations of rank-deficient blocks, while the resulting set of \( \mathcal{H} \)-matrix equations is solved using the generalized minimum residual (GMRES) iterative algorithm.

It should be remarked that the proposed approach closely resembles the method developed in ref \(^ {65} \). In this work characteristic modes of the elements in the reflectarray are used as macro base functions in MoM, and then the dominant characteristic modes are reused to reduce the order of MoM matrix. The overall procedure of the proposed simulation approach and the role of different techniques are shown in a flow chart in Figure 1. The details regarding the formulation and implementation are rather technical and are reported in the Supporting Information.
To demonstrate the applicability of the procedure outlined above, this section is organized as follows. First, we consider a one-atom-thick metasurface made of a graphene layer with graded circular chemical potential interruptions. The electrically small interruptions can be modeled by electric dipoles with 2D isotropic polarizabilities, which can be obtained analytically by considering a quasi-static response. Next, we apply the method to model an ultrathin plasmonic metasurface made of V-shaped nanoantennas. The plasmonic nanoantennas exhibit a very weak artificial magnetic response; however, bianisotropic effects may arise because of chirality of the elements. This is taken into account by using a bianisotropic polarizability tensor. Finally, the approach is used to simulate dielectric metasurfaces made of silicon nanobars. In this case, there is a strong magnetic response due to the high-index dielectric material, and the elements are modeled by anisotropic polarizability tensors. The solver is developed in MATLAB, and simulations are performed on a Linux-based desktop machine with Intel Xeon(R) processors 2.30 GHz for 251.8 GB of RAM utilizing 24 threads in parallel pools.

**One-Atom-Thick Metasurface.** Graphene has emerged as a promising alternative to plasmonic metals. It is shown to support deep-subwavelength surface plasmons at terahertz frequencies with considerably lower energy loss and stronger field localization. The chemical potential of graphene can be locally tuned by applying local electrostatic gating or chemical doping. The local interruption of the chemical potential in an otherwise perfect infinite graphene layer can be regarded as a surface inclusion embedded in the graphene background, which opens the possibility to realize any desired gradient index profile and thus brings a large flexibility to control and manipulate the propagation of surface plasmons.

In ref 68 by using an equivalent problem formulation, the polarizability of an electrically small circular chemical interruption has been derived that enables one to model the interruption as an electric dipole over the graphene layer. This allows for modeling the response of metasurfaces made of graded interruptions using the free-space Green’s function in the presence of a graphene sheet. The quasi-static approximation holds true as long as the spacing between the circular interruptions is \( t > 3D \), where \( D \) is the diameter of the circular interruption. In this case, there is no need for retrieval of the polarizability tensor. The unit cell of the graphene metasurface, as shown in Figure 2a, consists of a circular interruption with a different chemical potential (green region) in a graphene layer placed on a silica substrate. The unit cell size must be smaller than the plasmonic wavelength so that it can be modeled with a dipole moment and also can be viewed as homogeneous in order to realize an effective refractive index for the surface wave. The polarizability of the interruption is given by:

\[
\alpha_{\text{eff}} = \frac{-i\omega D^2}{4} \left[ 2\sigma_1 \sigma_2 + \frac{k_p D}{\omega \varepsilon_0} \left( \sigma_2 - \sigma_1 \right) \right] \frac{1}{\sigma_2 \sigma_1 + \sigma_2} \frac{\sigma_2 - \sigma_1}{\sigma_2 \sigma_1 + \sigma_2}
\]

(7)

Considering a layer of uniformly doped graphene placed on a silica half-space, the dispersion relation of surface plasmon polaritons (SPPs) at the interface of silica–air can be calculated in the nonretarded region \( (k_p \gg \alpha \sqrt{\mu \varepsilon_0}) \) from

\[
k_y = \frac{i \omega \varepsilon_0 \left( \varepsilon_{y1} + \varepsilon_{y2} \right)}{\sigma}
\]

(8)

where the conductivity of graphene \( \sigma \) is obtained from the Kubo formula. \( \varepsilon_{y1} \) and \( \varepsilon_{y2} \) are the dielectric constants of the materials above and under the graphene layer, which in this case correspond to air and silica, respectively. \( k_y \) is the surface wave vector, and the propagation direction is taken along \( y \) with the graphene lying in the \( x-y \) plane.

The effective refractive index for the SPPs can be obtained from the dispersion relation as \( n_{\text{eff}} = c_0 / \omega \), which is the relative scale of the plasmonic wavelength and the free-space wavelength. In order to design a metasurface with a gradient effective refractive index profile, we use the effective medium theory (EMT). According to the EMT, the effective refractive index of the unit cell shown in Figure 2a for a TM-polarized wave can be expressed as:

\[
N_{\text{eff}} = (1 - FF)n_{\text{eff} - 1} + FFn_{\text{eff} - 2}
\]

(9)

where FF = \( \pi D^2 / P^2 \) is the filling factor of the unit cell. \( N_{\text{eff}} \) is shown in Figure 2b as a function of filling factor, when \( \sigma_1 \) and \( \sigma_2 \) correspond to the graphene with doping levels of \( \mu_{c1} = 0.16 \) eV and \( \mu_{c2} = 0.4 \) eV, respectively, with a relaxation time of \( \tau = 0.5 \) ps.

Here, we use graded patterns of such chemical potential interruptions to design a system of four Luneburg lenses with touching rims that can mimic the cloaking devices in bulk medium. A Luneburg lens focuses a planar wavefront to a point image at the rim of the lens and vice versa. The refractive index profile of the lens is given by:

\[
N_{\text{eff}}(y, x) = \text{constant}
\]

where the constant is determined by the filling factor.
design at the excitation frequency of 30 THz, where the radius of each Luneburg lens is \( R = \lambda_{\text{spp}} = 97.5 \text{ nm} \) and the unit-cell period is fixed at \( \lambda_{\text{spp}} / 10 \). Note that since the maximum required filling ratio is 0.5, all the surface inclusions satisfy the quasi-static condition, which assures that they can be accurately modeled with the given polarizability in the quasi-static limit.\(^6\)

We use the dipolar equations to simulate the whole structure with a surface wave launched in the \( y \)-direction with \( K_{\text{spp}} \) corresponding to the SPP at the excitation frequency of 30 THz. The total number of inclusions is \( N = 1264 \), which leads to a problem of 2528 unknowns using isotropic 2D polarizability tensors in the proposed simulation approach. The problem can be solved within a few minutes thanks to the developed solver, which is a significant computational gain compared to standard numerical methods taking a long time because of the fine meshing required for discretizing the graphene layer.

Figure 3c and d show the real part and intensity of the field in the \( x \)-\( y \) plane, respectively. When a planar SPP wave is launched from the left side, the output remains planar while going around the central area, which makes it practically invisible. According to the transformation optics, the design maps all the rays inside the lenses so that the intensity in the area between the lenses (with a surface area of \( (4 - R)R^2 \)) is minimum. To demonstrate it more clearly, the intensity distribution along the white dashed line marked in Figure 3d is illustrated in Figure 3e. Such an invisible area may be suitable for hiding an electrode in photonic integrated circuits without disturbing the functionality of the device.\(^6\)

**Plasmonic Metalens.** Next, we apply the method for modeling a more complex element whose shape does not allow an analytical solution for the polarizability. We consider an ultrathin plasmonic metasurface consisting of V-shaped nanoantennas with various arm lengths and opening angles. These nanoantennas can be used to modulate the phase of the cross-polarized wave. When a linearly polarized electromagnetic wave is illuminated on the nanoantenna unit, two "symmetric" and "antisymmetric" modes can be excited. According to the theory proposed by Capasso, when the cross-polarized field is scattered by these two modes, its amplitude and phase can be modulated by changing the geometrical parameters of the nanoantenna.\(^7\) Figure 4a illustrates the schematic of a gold V-shaped nanoantenna that is placed on the surface of a silicon substrate with a thickness of \( t = 390 \text{ nm} \). The height (\( h \)) and width (\( w \)) of the V-shaped antenna are chosen as \( h = 50 \text{ nm} \) and \( w = 75 \text{ nm} \), respectively. The dashed line indicates the symmetry axis of the V-shaped nanoantenna. The angle between the polarization of incident light (which is along the \( y \)-axis) and symmetry axis is denoted by \( \beta \). This angle is set as \( \beta = 45^\circ \) so that both symmetric and antisymmetric modes can be

\[
N_{\text{eff}}(y, x) = N_{\text{eff}}(0) \sqrt{2 - \left( \frac{r}{R} \right)^2}
\]

where \( N_{\text{eff}}(0) \) is the effective refractive index at the center in the \( x \)-direction, \( r \) is the radial distance from the center of the lens, and \( R \) is the radius of the lens.

The index profile and the corresponding filling factors in the \( y \)-\( x \) plane are shown in Figure 3a and b, respectively for the geometry of graded chemical interruptions in each Luneburg lens is corresponding to the SPP at the excitation frequency of 30 THz, where the radius \( R \) is fixed at \( \lambda_{\text{spp}} / 10 \). Note that since the maximum required filling ratio is 0.5, all the surface inclusions satisfy the quasi-static condition, which assures that they can be accurately modeled with the given polarizability in the quasi-static limit.\(^6\)

![Figure 3](image.png)

**Figure 3.** (a) Distribution of the effective refractive index and (b) the corresponding filling factor of the metasurface unit cells in the \( y \)-\( x \) plane. (c) Real part and (d) intensity of the field in the \( y \)-\( x \) plane. The geometry of graded charged interruptions in each Luneburg lens is illustrated in the inset of (c). (e) Intensity distribution along the white dashed line in (d).

![Figure 4](image.png)

**Figure 4.** (a) Schematic of the metasurface unit cell consisting of a plasmonic V-shaped nanoantenna placed on a silicon substrate. (b) Cross-polarized transmission of the nanoantenna as a function of opening angle and arm length. (c and d) Normalized phase delay of the transmitted cross-polarized wave for the V-shaped nanoantenna and its mirror structure, respectively. The insets in the figures show the incident and scattered field components with respect to the symmetry axis of the structures. (e and f) Quantized transmission phase and amplitude corresponding to each antenna unit.
excited, contributing to the cross-polarized scattered wave. A y-polarized plane wave with an operating frequency of 193.54 THz (corresponding to the telecommunication wavelength of 1550 μm) is impinged on the structure from the top. Periodic Green’s functions are used to model the unit cells via DDA, and the effect of the substrate is taken into account using complex images. The mesh-size used in DDA for discretization of the elements is chosen as 5 nm to minimize the shape reconstruction errors due to the cubic lattice and ensure convergence. The period of the unit cell is fixed at p = 400 nm along both the x- and y-directions. The refractive index of the silicon substrate is 3.4 in this frequency. The amplitude and phase modulation of a cross-polarized transmitted wave are shown in Figure 4b and c, versus the opening angle and the arm length of the nanoantenna. As it can be seen, the transmission amplitude changes from 0 to 0.4 and the phase shift varies from 0° to 190°. In order to achieve a full phase shift coverage from 0° to 360°, one can use the mirror structure of the V-shaped nanoantenna by rotating the nanoantenna by 90° (i.e., β = 135°). In this case, the amplitude of the cross-polarized transmitted light remains the same, while an additional 180° phase shift is introduced in the mirror structure, which is shown in Figure 4d.

In order to realize a gradient phase profile ranging from 0° to 360° for the cross-polarized transmission while maintaining a uniform transmitted amplitude, 10 nanoantenna units are selected within the dashed contour shown in Figure 4b. The phase of the cross-polarized transmission corresponding to antenna unit 1 is defined as the reference. Figure 4e and f show the quantized amplitude and phase modulation obtained with the selected antenna units. The antenna unit parameters are shown in Table 1. It can be observed that the cross-polarized transmitted amplitudes do not have a strong variation and there is a constant phase difference of 36° among the 10 nanoantennas.

Table 1. Geometric Parameters of the Selected Nanoantennas for Quantization of Amplitude and Phase of the Cross-Polarized Transmitted Wave

<table>
<thead>
<tr>
<th>antenna unit</th>
<th>arm length (nm)</th>
<th>opening angle (deg)</th>
<th>β (deg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>220</td>
<td>10</td>
<td>45</td>
</tr>
<tr>
<td>2</td>
<td>178.9</td>
<td>60</td>
<td>45</td>
</tr>
<tr>
<td>3</td>
<td>137.9</td>
<td>30</td>
<td>45</td>
</tr>
<tr>
<td>4</td>
<td>124.2</td>
<td>60</td>
<td>45</td>
</tr>
<tr>
<td>5</td>
<td>96.84</td>
<td>70</td>
<td>45</td>
</tr>
<tr>
<td>6</td>
<td>213.2</td>
<td>40</td>
<td>135</td>
</tr>
<tr>
<td>7</td>
<td>151.6</td>
<td>50</td>
<td>135</td>
</tr>
<tr>
<td>8</td>
<td>137.9</td>
<td>80</td>
<td>135</td>
</tr>
<tr>
<td>9</td>
<td>137.9</td>
<td>120</td>
<td>135</td>
</tr>
<tr>
<td>10</td>
<td>144.7</td>
<td>180</td>
<td>135</td>
</tr>
</tbody>
</table>

Dielectric Meta-Hologram. We next apply the method to modeling dielectric metasurfaces. We consider silicon nanobars of different sizes to realize a multicolor hologram. Recently, metasurface holograms have become a hot topic of research. Several research efforts have been conducted to construct metasurfaces having multicolor responses in the reflection. Here, we aim at a theoretical demonstration of a transmissive multicolor phase-only hologram using three different unit cells responding to three primary colors. As it has been shown in ref 19, the primary color unit cells can be multiplexed into a single metasurface to reconstruct arbitrary RGB images, leading to a full-color hologram.

The aim in a full-color hologram is to illuminate the metasurface with white light and get the reconstructions of the three RGB components of a target image at the same spots with the same sizes on a specified target plane. Figure 7a shows this working principle; the metasurface is illuminated by a white light incidence and imprints the phases for each color component separately through corresponding subcells. For this objective, one needs three different subcells operating at the wavelength of the three primary colors with a relatively narrow spectral response, which is essential to alleviate the color crosstalk for colorful holograms. Moreover, the elements should be able to imprint the required phase maps for the hologram. In order to meet these requirements, we...
utilize silicon nanobars to imprint the Pancharatnam–Berry (PB) phase because of relatively low loss in the visible range compared with plasmonic nanoantennas and narrow spectral response for low crosstalk between different colors. The phase modulation mechanism is based on the PB shift for a circularly polarized incident wave. By clockwise rotation of the silicon nanobars with an angle denoted as $\theta$ around the geometric center, a part of the incident circularly polarized light is converted to a beam of opposite helicity and picks up a geometric phase equal to $2\theta$. This part corresponds to the transmitted circular polarization with opposite handedness and the reflected circular polarization with the same handedness. This orientation-controlled phase covers the range of 0° to 360°, thereby providing full control for the wavefront manipulation.

We tune the geometric parameters of the silicon nanobars on a glass substrate with a thickness of 250 nm to obtain three unit cells operating at the wavelength of the three primary colors with a relatively narrow spectral response. The permittivity of glass and silicon are extracted from the experimental data.

Figure 7b shows the obtained unit cells. The length, width, height, and period of the designed unit cells are blue {125 nm, 35 nm, 120 nm, 150 nm}, green {250 nm, 60 nm, 80 nm, 300 nm}, and red {335 nm, 80 nm, 120 nm, 390 nm}.

The unit cells are simulated using DDA with periodic Green’s functions. The mesh size is chosen as 5 nm to accurately reconstruct the nanobars using a cubic lattice and ensure convergence. The cross-transmission for circularly polarized light is obtained using Jones calculus as

$$t_{LR} = \frac{(t_{xx} - t_{yy} - (t_{xx} + t_{yy}))}{2} \quad (11)$$

where $t_{xx}$, $t_{yy}$, $t_{xy}$, and $t_{yx}$ are the reflection coefficients for the linearly polarized light.

Figure 7c demonstrates the transmittance spectra of the left-handed circularly polarized light for each element upon the incidence of right-handed circularly polarized light. The operating wavelength of each element is blue $\{\lambda_b = 459\text{ nm}\}$, green $\{\lambda_g = 512\text{ nm}\}$, and red $\{\lambda_r = 650\text{ nm}\}$. One can notice that the transmittance of the nanobar for blue light is lower than that for green and red, which is attributed to the increasing loss of silicon toward the short wavelength.

The nanobars can be rotated to pick up a PB phase shift while still maintaining a narrow spectral response. The phase delay and amplitude variations for each unit cell corresponding
to RGB colors are shown in Figure 7d and e, respectively. We use nine elements with a rotation spacing of 20 degrees to realize a full phase range of 0° to 360° (shown by markers in Figure 7d and e).

Even though the nanorods are strictly achiral, unlike the plasmonic elements they exhibit a strong artificial magnetic response. As such, one would need an anisotropic 6 × 6 polarizability tensor to describe each element. Six independent measurements of induced electric and magnetic dipole moments are performed to retrieve the polarizability. It should be noted that there is no need for retrieving the polarizability tensor for each nanoantenna unit. The polarizability tensor is retrieved only once for each nanobar and then is rotated with the corresponding rotation angle as

\[
\hat{a}_\theta = \begin{bmatrix}
\hat{R}_\theta & 0 \\
0 & \hat{R}_\theta
\end{bmatrix}
\hat{a}
\begin{bmatrix}
\hat{R}_\theta & 0 \\
0 & \hat{R}_\theta
\end{bmatrix}
\tag{12}
\]

where \( \hat{R}_\theta = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \) and \( \dagger \) denotes the transposed matrix.

The data of the retrieved polarizability components of the elements are shown in the Supporting Information.

We aim at reconstructing the blue, green, and red components of the "Northeastern" logo using three metasurfaces with corresponding color subcells at a focal distance of \( 5\lambda \), where \( \lambda \) is the operating wavelength of each subcell. The resolution of the target image is set such that the pixel sizes are less than half the operating wavelength to satisfy the Shannon–Nyquist sampling theorem. The size of each metasurface is \( N = 700 \times 190 \) elements, which is equivalent to a problem of \( 6N = 798k \) unknowns using the simulation approach offered. The simulation time for each metasurface with the developed solver is around 3 h. It should be noted that the structure seems impossible to solve on standard computational models.

Figure 8a shows the section of the blue component hologram corresponding to the letter "N" with an enlarged view and meshing of the unit cells in DDA in a schematic representation of the metasurface. The discretized required phase distribution in the metasurface plane is obtained using the computer holography technique and is demonstrated in Figure 8b. Figure 8c–e show the reconstructed holographic images corresponding to the primary colors blue, green, and red, respectively.

As it can be seen, the target image is well reconstructed using the metasurface hologram. A speckle pattern is observed in all the reconstructed images due to the phase discreteness, which can be further improved by increasing the levels of phase quantization. One can notice that the performance deteriorates as we move from blue to red. This is due to a stronger undesirable variation in the amplitude upon rotation of the silicon nanobars (Figure 7e). It is also important to note that the design presented here is a phase-only hologram.
without amplitude modulation. As it has been shown in monochromatic hologram designs, an amplitude modulation improves the performance of the hologram. However, obtaining amplitude modulation in the designs based on PB phase shift is not possible, as there is no independent control over the amplitude and phase of the transmission.

**COMPUTATIONAL LOAD AND ACCURACY**

In the previous section, we presented that the dipolar analysis can be applied to an aperiodic array of subwavelength elements to handle reduced order analysis with a significant efficiency while correctly predicting the behavior of metasurfaces. In this section, we evaluate the performance of the proposed method regarding the accuracy and efficiency through comparison of the results with full-wave simulations.

In order to evaluate the amount of error introduced by ignoring the higher order modes, first, we look at the extinction cross section of the elements and compare the contributions of dipolar and higher order modes to the extinction. Figure 9 shows the results for the V-shaped nanoantenna units and the silicon nanobars corresponding to blue, green, and red subcells of the metasurface corresponding to the blue hologram. The normalized amplitudes of the cross-transmitted circularly polarized electric field $E_{1B}$ obtained with (b) the proposed method and (c) FDTD solver and (d) their absolute difference. The phases of the transmitted cross-polarized circularly polarized electric field $E_{1B}$ obtained with (e) the proposed method and (f) FDTD solver and (g) their absolute difference.

The results for the normalized amplitudes and phases of the cross-transmitted circularly polarized electric field $E_{1B}$ as well as the corresponding absolute differences between the results of both methods are shown in Figure 10. A remarkable consistency is observed between the field patterns in both phase and amplitude. It can be seen that the absolute difference between the amplitudes is less than 0.11. The average absolute differences of the amplitude and phase on the whole grid are 0.02 and 7 degrees, respectively. The differences observed should mainly result from numerical errors inherent in the FDTD simulation and the approximations made in the proposed method. Additionally, in the FDTD method the field components are arranged in a staggered grid, while in the
calculation of the proposed method this arrangement has not been regarded through simply rotating the polarizability tensors, implying that the location of a certain field component might be slightly shifted compared to the location of the corresponding component in the FDTD grid. Improvements could be made to further increase the consistency, including increasing the mesh size in FDTD and experimenting with different retrieval conditions for the polarizabilities in the proposed method.

The computation times and memory requirements for the proposed method and FDTD solver are compared in Table 2. The duration for the proposed method includes the polarizability retrieval time, matrix fill time (self-interaction and coupling terms), and the matrix solution time. Since we are dealing with circular polarization, we require the transmission coefficients for two orthogonal linear polarizations. As such, the FDTD simulation is performed twice, while in the proposed method, the interaction matrix is built only once and the matrix solution is obtained twice with two different excitation vectors. Drastic reductions have been observed in the computation time by a speed up factor of more than 23 times and the memory requirement by more than 31 times, which demonstrate the superior efficiency of the proposed method.

| Table 2: Comparison of the Computation Times and Memory Requirements for the Proposed Method and FDTD Solver |
|---------------------------------|------------------------|-----------------|
| number of unknowns  | proposed method | FDTD   |
| memory requirement  | 4704         | 5.04 × 10^{11} |
| solution time  | polarizability retrieval | 354.2 MB | 11.1 GB |
|                  | self-terms | 2.5 s | 2h=4h |
|                  | coupling terms | 3.4 min | \sim 2 \times 2 h = 4 h |
|                  | matrix solution | 2 × 7.5 s |
|                  | total | \sim 10.1 min |

The efficiency of the method is best appreciated when it is intended to analyze the effects of changes in the arrangement of the elements. Once the polarizabilities are retrieved and tabulated, only the self-interaction terms need to be recalculated, and the computational time for solving the resulting matrix equations is 15 s. This paves the way for control and optimization of the metasurfaces. The significant efficiency afforded by the proposed method can result in saving days in the optimization procedure. It also decreases the control space of the optimization problem by decreasing the number of unknowns. As the method does not require the computationally expensive preprocessing, it leaves much room for optimization to obtain desired characteristics through a reverse engineering procedure.

To test the limits of the proposed approximation and estimate the tolerance of the method versus higher order mode excitations, we consider a 26 × 26 element red meta-hologram reconstructing the letter “n” at a focal distance of 5λ. The phase distribution over the elements of the meta-hologram is shown in Figure 11a. A mesh size of Δx = Δy = 0.0323λ and Δz = 0.0185λ is used for the FDTD method to minimize the reconstruction errors in the staggered FDTD grid. The field patterns are obtained at the target plane with a similar average absolute differences of the amplitude and phase on the whole grid are 0.07 and 20 degrees, respectively. It should be remarked that stronger excitations of higher order modes are rarely encountered in the ultrathin metasurfaces due to the small effective size of the elements, and thus it is expected that the proposed method will attain reasonable accuracy within the given levels of deviation from numerical methods. It should also be noted that in many cases the higher order modes are avoided in the design of the elements due to the undesirable effects on the performance.

It is important to note that what we are looking for is a balance between speed and accuracy. In the cases where the thickness of the elements becomes comparable to the wavelength stronger higher order modes may be excited. In these cases, the proposed simulation approach can be extended to take into account the role of higher order modes by retrieving a T-matrix of the elements and extending the modal analysis to higher order terms. In this case, instead of dipolar equations one is dealing with the equations for vector spherical harmonics. This approach can increase the accuracy while still yielding the order reduction because of the few number of modes excited in each element. The challenge in this implementation will be developing acceleration techniques for calculation of Weyl-type integrals, which characterize the effect of substrate scattering for higher order modes, and reducing the computational complexity of the system of equations in which interaction between the elements is described with the translation coefficients instead of Green’s functions. Addressing
these challenges should be the solution to the ultimate efficiency in treating large-scale aperiodic arrays of subwavelength elements up to the desired accuracy, which remains a topic of future research. In the case where higher order modes are negligible, dynamic polarizability can be regarded as equivalent to the T-matrix of the element.

**CONCLUSION**

In this paper, a powerful simulation approach is proposed, in order to manage reduced order analyses of large-scale metasurfaces based on a hierarchical dipole approximation. The proposed technique has been framed within the DDA formalism incorporating acceleration techniques of DCIM and the H-matrix, which allows the ultraefficient simulation of large-scale metasurfaces of arbitrary elements and to solve structures that current numerical models are unable to tackle. Several examples of the applicability of the method have been presented for modeling one-atom-thick, plasmonic and dielectric metasurfaces. The method has allowed attaining a significant computational gain compared to the full-wave simulations without affecting the accuracy of the results and comprehensively predicting the functionality of the metasurfaces owing to the robust consideration of couplings. As such, the method paves the way for the optimization of the metasurfaces. Moreover, thanks to the DDA component, the method can be extended to analyze advanced nanophotonics devices such as molecular-scale sensors and quantum dot platforms.

**ASSOCIATED CONTENT**

1. Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsphto-000568.

- Dyadic Green’s functions; periodic Green’s functions: Ewald acceleration method; discrete complex image method; hierarchical matrix construction; retrieved polarization data of the elements (PDF)

**AUTHOR INFORMATION**

**Corresponding Author**

*E-mail: hosseinnm@ece.neu.edu.*

**ORCID**

Mohammad Mahdi Salary: 0000-0003-2056-3348

**Notes**

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

This work is supported in part by the Defense Advanced Research Projects Agency (DARPA) (via grant number N00014-14-1-0850) and in part by the U.S. Air Force Office of Scientific Research (AFOSR), #FA9550-14-1-0349.

**REFERENCES**


