Computing parametrized solutions for plasmonic nanogap structures

F. Vidal-Codina *, N.C. Nguyen, J. Peraire

Dept. of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

A B S T R A C T

The interaction of electromagnetic waves with metallic nanostructures generates resonant oscillations of the conduction-band electrons at the metal surface. These resonances can lead to large enhancements of the incident field and to the confinement of light to small regions, typically several orders of magnitude smaller than the incident wavelength. The accurate prediction of these resonances entails several challenges. Small geometric variations in the plasmonic structure may lead to large variations in the electromagnetic field responses. Furthermore, the material parameters that characterize the optical behavior of metals at the nanoscale need to be determined experimentally and are consequently subject to measurement errors. It then becomes essential that any predictive tool for the simulation and design of plasmonic structures accounts for fabrication tolerances and measurement uncertainties.

In this paper, we develop a reduced order modeling framework that is capable of real-time accurate electromagnetic responses of plasmonic nanogap structures for a wide range of geometry and material parameters. The main ingredients of the proposed method are: (i) the hybridizable discontinuous Galerkin method to numerically solve the equations governing electromagnetic wave propagation in dielectric and metallic media, (ii) a reference domain formulation of the time-harmonic Maxwell's equations to account for arbitrary geometry variations; and (iii) proper orthogonal decomposition and empirical interpolation techniques to construct an efficient reduced model. To demonstrate effectiveness of the models developed, we analyze geometry sensitivities and explore optimal designs of a 3D periodic coaxial nanogap structure.

© 2018 Elsevier Inc. All rights reserved.

The field of plasmonics [30,40] studies the excitation of electrons in the conduction band of metallic nanostructures. This excitations can lead to resonances which result in the trapping of light in deep-subwavelength regions, leading to enormous near-field enhancements of the incident electromagnetic (EM) field. These regions of high-intensity EM field are typically observed near the corners, sharp features or cavities within metallic nanostructures. Moreover, the tight confinement and enhancement of the EM waves provide unique means for manipulating light and its interaction with metals, well beyond the diffraction limit. As a result, the development of plasmonic field enhancement structures has found applications in energy harvesting [5], sensing [54], plasmonic waveguiding [31] and near-field scanning microscopy [38].

---

* Corresponding author.
  E-mail addresses: fvidal@mit.edu (F. Vidal-Codina), cuongng@mit.edu (N.C. Nguyen), peraire@mit.edu (J. Peraire).

The ability to accurately model and simulate EM wave propagation for plasmonics requires capabilities that challenge traditional simulation techniques. Plasmonic phenomena involve the propagation of μm and mm wavelengths through nanoscale apertures. It is thus essential to devise simulation methods that are capable of resolving multiple length scales. From a manufacturing perspective, the fabrication of nanometer-wide gap structures is challenging, and minor geometric deviations from the nominal design may lead to significant alterations in the performance of the device. Furthermore, the optical behavior of the metals at the nanoscale is poorly characterized and the material constants in the models, which are determined experimentally, are subject to measurement errors. If our ultimate goal is to develop simulation methods that can be effectively used in the design of plasmonic structures, it is of foremost importance that these methods incorporate fabrication tolerances as well as parameter uncertainties.

The goal of this paper is to combine an advanced numerical simulation method for the time-harmonic Maxwell's equations, the hybridizable discontinuous Galerkin (HDG) method, with a reduced order model (ROM) framework that will enable us to efficiently compute solutions for different combinations of the input parameters. These parameters may include material properties, frequency and direction of the incident light, as well as geometric dimensions. The ability to inexpensively obtain multiple solutions for different input parameters is critical to allow the effective exploration of the design space. The HDG method [11,33], has been successfully applied to acoustics and elastodynamics [34,50], as well as electromagnetics [35]. The HDG method is fully implicit in time, can be implemented on unstructured spatial discretizations, and can be made high-order accurate in both space and time, thus exhibiting low dissipation and dispersion. Compared to other DG methods [26], the HDG method results in a reduced number of globally coupled unknowns, thereby resulting in smaller linear systems, and exhibits optimal convergence rates for both the solution and the flux. Despite the computational advantage offered by the HDG method, the size of the resulting systems of equations, typically several hundred thousand, is still too large to allow for the large number of solutions required in the design process. For this reason, we propose the construction of reduced order models (ROM) built upon the HDG method. These ROM allow for rapid and inexpensive queries of the outputs of interest for a range of geometry and problem parameters, effectively providing a parametrized solution of the problem.

The ROM of interest here are created by constructing a low-dimensional basis of the solution state space and seeking approximate solutions within this reduced space via Galerkin projection or other means. For this reason, they are often referred to as reduced basis (RB) methods [24,46,47]. Proper orthogonal decomposition (POD) [51] is commonly used to construct such low-dimensional bases. The starting point for POD is a set of solutions, or snapshots, that are obtained by numerically solving the forward HDG model for different values of the input parameters. Singular value decomposition is then employed to determine a low-dimensional basis such that the dominant information is retained by the basis vectors. POD has been used to construct ROM in computational fluid dynamics [4], mechanical systems [27] and electromagnetics [1]. In addition, alternative approaches to POD for creating ROM have been proposed. For instance, Arnoldi methods can be used to develop ROM for turbomachinery flows [58] and in recent years, Patera et al. [19,45,49] introduced greedy sampling strategies to adaptively construct the RB, as well as rigorous *a posteriori* error estimators to certify the quality of the approximation and an offline/online strategy to improve computational efficiency. Other work has also been conducted to incorporate *a priori* error analysis [17,29] of the approximation in ROM. The reduced RB representation can also be used to compute rapid and accurate approximations of functional outputs of parametrized PDEs [37].

Recently, RB methods have been used for the time-harmonic Maxwell's equations in combination with DG [10,13,28, 44] and integral [16,25] methods to successfully generate fast responses for complex electromagnetic devices, where the frequency and material properties are treated as parameters. Furthermore, there has been considerable work to incorporate geometry as a parameter for electromagnetic applications, using either affine geometry transformations [9,18,23,57] or parametrized meshes [22].

The application of ROM techniques to linear PDEs with affine parameter dependence is fairly straightforward. However, the existence of nonlinearities and more complex parametric dependence, such as geometry parametrization, poses a severe challenge for RB. In that case, the projection required to construct the reduced order model cannot be pre-computed and its cost scales as the dimension of the original solution space. In order to overcome this limitation, the empirical interpolation method (EIM) [2] and its discrete counterpart DEIM [8] were proposed. These methods employ a greedy sampling strategy to generate an independent approximation to the nonlinear terms, consisting of a weighted combination of interpolation functions. Thus, replacing the nonlinear term with this interpolation approximation recovers an efficient computational strategy to perform the projection onto the low-dimensional space. In electromagnetics, the EIM has been leveraged to construct a ROM for 3D scattering problems [18]. Alternatively, geometry parametrization has also been considered by directly interpolating the projection matrices that arise from the spatial discretization [15], with applications to 2D and 3D waveguides [6,7].

In order to pursue ROM within the HDG method, we first develop a reference domain formulation [43] that renders a transformed set of time-harmonic Maxwell's equations for an arbitrary differentiable deformation mapping. We then combine the HDG discretization of the transformed equations with empirical interpolation techniques to recover a weak HDG formulation in terms of affine parametric expressions [49]. Finally, we employ POD to construct the RB for the ROM and the standard offline-online computational strategy that enables the evaluation of approximate solutions of large 3D problems in real-time. The main contribution of this article is the combination of existing numerical approaches such as HDG, EIM and ROM to address the simulation and design of novel plasmonic devices. The framework presented here
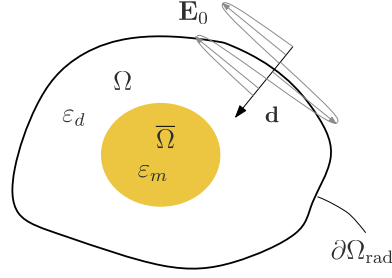
Fig. 1. Metallic structure $\overline{\Omega}$ embedded in dielectric Ω illuminated by plane wave.

is shown to dramatically reduce the computational cost of plasmonic simulations, therefore enabling the exploration of multiple material and geometry configurations.

This article is organized as follows. In Section 1, we introduce the equations and notation used throughout the paper. In Section 2, we introduce the HDG method to solve time-harmonic Maxwell's equations on a reference domain and discuss its implementation. In Section 3, we describe the construction of a reduced order model for time-harmonic Maxwell's equations based on proper orthogonal decomposition and empirical interpolation. In Section 4, we present numerical results to assess the performance of the proposed method. We finalize in Section 5 by providing some concluding remarks.

1. Governing equations

1.1. Time-harmonic Maxwell's equations

Maxwell's equations describe the propagation of the electric $\mathcal{E}(\mathbf{x}, t)$ and magnetic $\mathcal{H}(\mathbf{x}, t)$ fields. For monochromatic waves, we can express the electromagnetic fields using only their complex amplitudes, since the time dependence is assumed harmonic, that is $\mathcal{E}(\mathbf{x}, t) = \Re\{\mathbf{E}(\mathbf{x}) \exp(-i\omega t)\}$, for a given angular frequency ω and corresponding frequency $f = \omega/(2\pi)$. For a linear, non-magnetic medium in the absence of impressed currents and charges, the frequency-domain Maxwell's equations for the complex electromagnetic amplitudes $\mathbf{E}$ and $\mathbf{H}$ are expressed as

$$\nabla \times \mathbf{E} - i\omega\mathbf{H} = 0,$$
$$\nabla \times \mathbf{H} + i\omega\varepsilon\mathbf{E} = \sigma\mathbf{E},$$

where ε is the relative permittivity of the medium and σ is its conductivity.

In this paper, we study the scenario where a plasmonic structure $\overline{\Omega}$ embedded in a dielectric medium Ω scatters an incident **p**-polarized plane wave $\mathbf{E}_0$ propagating in the **d**-direction, that is $\mathbf{E}_0 = \mathbf{p}\exp(i\omega\sqrt{\varepsilon_d}\,\mathbf{d}\cdot\mathbf{x})$, see Fig. 1. The dielectric medium is characterized by a dielectric permittivity ε_d and zero conductivity. The metal $\overline{\Omega}$ can be described with the complex Drude's permittivity [14], that is

$$\varepsilon_m = \varepsilon_\infty - \frac{\sigma}{i\omega} = \varepsilon_\infty - \frac{\omega_p^2}{\omega(\omega + i\gamma)}$$

where ε_∞ is permittivity due to the valence-band bound electrons, ω_p is the (nondimensional) plasma frequency of the metal and γ is the (nondimensional) electron collision rate. This model assumes the electrons in the valence band are fully detached from the ions, and are only subject to electron–ion and electron–electron collisions. This model accounts for the dispersive and lossy nature of metals and is commonly known in literature as local model, since the motion of one electron is not coupled to that of its neighbors. For compactness, we use a spatially-varying permittivity $\varepsilon(\mathbf{x})$, defined as

$$\varepsilon(\mathbf{x}) = \varepsilon_d \mathbb{1}_\Omega + \varepsilon_m \mathbb{1}_{\overline{\Omega}}.$$

Introducing the additional variable $\mathbf{V} = i\omega\mathbf{H}$, we arrive at the time-harmonic Maxwell's equations

$$\begin{aligned}
\nabla \times \mathbf{E} - \mathbf{V} &= 0, \\
\nabla \times \mathbf{V} - \omega^2\varepsilon(\mathbf{x})\mathbf{E} &= 0,
\end{aligned} \tag{1}$$

defined in $\overline{\Omega} \cup \Omega$. The system above is complemented with boundary conditions

$$\mathbf{n} \times \mathbf{E} \times \mathbf{n} = \mathbf{E}_\partial, \quad \text{on } \partial\Omega_E,$$
$$\mathbf{n} \times \mathbf{V} = \mathbf{n} \times \mathbf{V}_\partial, \quad \text{on } \partial\Omega_V,$$
$$\mathbf{n} \times \mathbf{V} + i\omega\sqrt{\varepsilon_d}\,\mathbf{n} \times \mathbf{E} \times \mathbf{n} = \mathbf{f}_0, \quad \text{on } \partial\Omega_{\text{rad}},$$
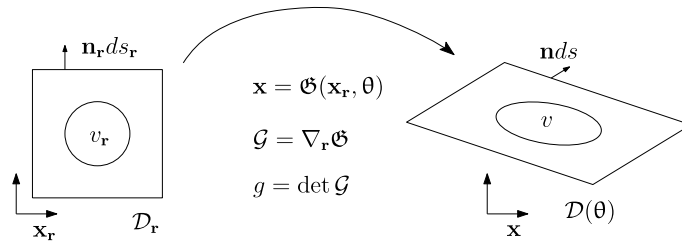
Fig. 2. Deformation mapping between reference and physical domain.

with $\mathbf{f}_0 = \mathbf{n} \times (\nabla \times \mathbf{E}_0) + i\omega\sqrt{\varepsilon_d}\mathbf{n} \times \mathbf{E}_0 \times \mathbf{n}$. The first (resp. second) equation corresponds to a boundary condition where the value of the tangential component of the electric (resp. magnetic) field is prescribed. Symmetries in the structure can be enforced by setting $\mathbf{E}_\partial = \mathbf{0}$ ($\mathbf{V}_\partial = \mathbf{0}$) on selected boundaries, see Section 4. The last equation imposes the radiation conditions required to truncate the infinite space. In particular, we employ Silver–Müller conditions, which are first order absorbing boundary conditions [32,53].

1.2. Reference domain formulation

In this section, we develop a reference domain formulation for the time-harmonic Maxwell's equations. Let us assume that the geometry of the physical domain, $\mathcal{D}(\theta) \in \mathbb{R}^n$ with coordinates $\mathbf{x}$, is parametrized by θ defined in a compact set. In addition, we consider a reference domain $\mathcal{D}_\mathbf{r} \in \mathbb{R}^n$ with coordinates $\mathbf{x}_\mathbf{r}$ which is mapped to the physical domain $\mathcal{D}(\theta)$ by means of a one-to-one mapping given by a diffeomorphism $\mathbf{x} = \mathfrak{G}(\mathbf{x}_\mathbf{r}, \theta)$, see Fig. 2. Our goal is to pull Maxwell's equations in the physical domain (1) back to the reference domain. The transformed equations in the reference domain are parametrized by θ, but in return, they are defined and can be solved on a fixed, reference space, thus avoiding re-discretizing the computational domain for each new geometry, or value of the mapping parameters θ.

The deformation gradient of the mapping $\mathfrak{G}$ and its Jacobian are defined $\mathcal{G} = \nabla_\mathbf{r}\mathfrak{G}$ and $g = \det \mathcal{G}$, respectively. Note that $\nabla = \mathcal{G}^{-T}\nabla_\mathbf{r}$. For simplicity, we shall assume deformations only occur within the vicinity of the metal, thus the outer boundaries of the computational domain – where radiation conditions are prescribed – remain fixed. For the plasmonic structures studied here, periodicity occurs in the $x - y$ directions and radiation conditions are imposed in the z constant plane, thus the mapping can be easily decomposed to satisfy the required conditions, see Appendix B. More general structures may either require more complex deformation mappings, or the use of perfectly matched layers (PMLs) [3] to model radiation without the constraint of keeping the radiation boundaries fixed.

In order to obtain the transformed equations on the reference space $\mathcal{D}_\mathbf{r}$, we consider, for illustration purposes, the second equation of (1) and integrate the curl term on a control volume $v \in \mathcal{D}$. The application of Stokes' theorem and the relations $dv = g dv_\mathbf{r}$ and $\mathbf{n} ds = g\mathcal{G}^{-T}\mathbf{n}_\mathbf{r} ds_\mathbf{r}$ [43] yields,

$$\int_v \nabla \times \mathbf{V} dv = -\int_S \mathbf{V} \times \mathbf{n} ds = -\int_{S_\mathbf{r}} g\mathbf{V} \times \left(\mathcal{G}^{-T}\mathbf{n}_\mathbf{r}\right) ds_\mathbf{r} = -\int_{S_\mathbf{r}} g\left(\mathcal{G}^{-T}\left(\mathcal{G}^T\mathbf{V}\right)\right) \times \left(\mathcal{G}^{-T}\mathbf{n}_\mathbf{r}\right) ds_\mathbf{r}$$

$$= -\int_{S_\mathbf{r}} \mathcal{G}\left(\mathcal{G}^T\mathbf{V}\right) \times \mathbf{n}_\mathbf{r} ds_\mathbf{r} = \int_{v_\mathbf{r}} \mathcal{G}\nabla_\mathbf{r} \times \left(\mathcal{G}^T\mathbf{V}\right) dv_\mathbf{r}.$$

The second term in transforms as

$$\int_v \omega^2 \varepsilon(\mathbf{x})\mathbf{E} dv = \int_{v_\mathbf{r}} \omega^2 \varepsilon(\mathbf{x}_\mathbf{r})\mathbf{E} g dv_\mathbf{r}.$$

The boundary conditions are transformed similarly, for instance

$$0 = \int_S (\mathbf{E} - \mathbf{E}_\partial) \times \mathbf{n} ds = \int_{S_\mathbf{r}} g(\mathbf{E} - \mathbf{E}_\partial) \times (\mathcal{G}^{-T}\mathbf{n}_\mathbf{r}) ds_\mathbf{r} = \int_{S_\mathbf{r}} \mathcal{G}\left[\mathbf{n}_\mathbf{r} \times (\mathcal{G}^T(\mathbf{E} - \mathbf{E}_\partial))\right] ds_\mathbf{r}$$

$$= \int_{S_\mathbf{r}} \mathbf{n}_\mathbf{r} \times (\mathcal{G}^T(\mathbf{E} - \mathbf{E}_\partial)) ds_\mathbf{r}.$$

Finally, the transformed time-harmonic Maxwell's equations in $\overline{\Omega}_\mathbf{r} \cup \Omega_\mathbf{r}$ read

$$\nabla_\mathbf{r} \times \boldsymbol{G}\mathbf{e} - \mathbf{v} = 0,$$
$$\nabla_\mathbf{r} \times \boldsymbol{G}\mathbf{v} - \omega^2 \varepsilon(\mathbf{x}_\mathbf{r})\mathbf{e} = 0,$$

$$(2)$$

where a generic field $\mathbf{u}$ transforms as $\mathbf{u} = g\mathcal{G}^{-1}\mathbf{U}$ and $\boldsymbol{G} = g^{-1}\mathcal{G}^T\mathcal{G}$. The system is complemented with boundary conditions

$$\mathbf{n_r} \times (\boldsymbol{G}\mathbf{e}) = \mathbf{n_r} \times (\boldsymbol{G}\mathbf{e}_\partial), \quad \text{on } \partial\Omega_{\mathbf{r},E},$$

$$\mathbf{n_r} \times (\boldsymbol{G}\mathbf{v}) = \mathbf{n_r} \times (\boldsymbol{G}\mathbf{v}_\partial), \quad \text{on } \partial\Omega_{\mathbf{r},V},$$

$$\mathbf{n_r} \times (\boldsymbol{G}\mathbf{v}) + i\omega\sqrt{\varepsilon}\,\mathbf{n_r} \times (\boldsymbol{G}\mathbf{e}) \times \mathbf{n_r} = \mathbf{f}_0, \quad \text{on } \partial\Omega_{\mathbf{r},\mathrm{rad}}.$$

Since deformations are only nonzero in the vicinity of the scatterer ($\mathfrak{G}\!\restriction_{\partial\Omega_\mathrm{rad}} = \mathrm{Id}$), we have $\mathbf{V} = \boldsymbol{G}\mathbf{v}$ and $\mathbf{E} = \boldsymbol{G}\mathbf{e}$ on $\partial\Omega_\mathrm{rad}$.

2. HDG method for time-harmonic Maxwell's equations

2.1. Approximation spaces

Following [35], we denote by $\mathcal{T}_h$ a triangulation of disjoint regular elements T that partition a domain $\mathcal{D} \in \mathbb{R}^3$. The set of element boundaries is then defined as $\partial\mathcal{T}_h := \{\partial T : T \in \mathcal{T}_h\}$. For an arbitrary element $T \in \mathcal{T}_h$, $F = \partial T \cap \partial\mathcal{D}$ is a boundary face if it has a nonzero 2D Lebesgue measure. Any pair of elements T^+ and T^- share an interior face $F = \partial T^+ \cap \partial T^-$ if its 2D Lebesgue measure is nonzero. We finally denote by $\mathcal{E}_h^o$ and $\mathcal{E}_h^\partial$ the set of interior and boundary faces respectively, and the total set of faces $\mathcal{E}_h = \mathcal{E}_h^o \cup \mathcal{E}_h^\partial$.

Let $\mathbf{n}^+$ and $\mathbf{n}^-$ be the outward-pointing unit normal vectors on the neighboring elements T^+, T^-, respectively. We further use $\mathbf{u}^\pm$ to denote the trace of $\mathbf{u}$ on F from the interior of $T^\pm$, where $\mathbf{u}$ resides in $\boldsymbol{L}^2(\mathcal{D}) \equiv [L^2(\mathcal{D})]^3$, with $L^2(\mathcal{D})$ the space of square integrable functions on $\mathcal{D}$. The jump $[\![\cdot]\!]$ for an interior face $F \in \mathcal{E}_h^o$ is defined as

$$[\![\mathbf{u} \odot \mathbf{n}]\!] = \mathbf{u}^+ \odot \mathbf{n}^+ + \mathbf{u}^- \odot \mathbf{n}^-,$$

and is single-valued for a boundary face $F \in \mathcal{E}_h^\partial$ with outward normal $\mathbf{n}$, that is

$$[\![\mathbf{u} \odot \mathbf{n}]\!] = \mathbf{u} \odot \mathbf{n},$$

where the binary operation $\odot$ refers to either $\cdot$ or $\times$. The tangential $\mathbf{u}^t$ and normal $\mathbf{u}^n$ components of $\mathbf{u}$, for which $\mathbf{u} = \mathbf{u}^t + \mathbf{u}^n$, are then represented as

$$\mathbf{u}^t := \mathbf{n} \times (\mathbf{u} \times \mathbf{n}), \qquad \mathbf{u}^n := \mathbf{n}(\mathbf{u} \cdot \mathbf{n}).$$

Let $\mathcal{P}^p(\mathcal{D})$ denote the space of complex-valued polynomials of degree at most p on $\mathcal{D}$. We now introduce the following approximation spaces

$$W_h = \{w \in L^2(\mathcal{D}) : w|_T \in \mathcal{P}^p(T), \ \forall T \in \mathcal{T}_h\},$$

$$\boldsymbol{W}_h = \{\boldsymbol{\xi} \in \boldsymbol{L}^2(\mathcal{D}) : \boldsymbol{\xi}|_T \in \left[\mathcal{P}^p(T)\right]^3, \ \forall T \in \mathcal{T}_h\},$$

$$M_h = \{\mu \in L^2(\mathcal{E}_h) : \mu|_F \in \mathcal{P}^p(F), \ \forall F \in \mathcal{E}_h\},$$

$$\boldsymbol{M}_h = \{\boldsymbol{\mu} \in \boldsymbol{L}^2(\mathcal{E}_h) : \boldsymbol{\mu}|_F \in \mathcal{P}^p(F)\mathbf{t}_1 \oplus \mathcal{P}^p(F)\mathbf{t}_2, \ \forall F \in \mathcal{E}_h\},$$

where $\mathbf{t}_1$, $\mathbf{t}_2$ are vectors tangent to the face, thus including the curl-conforming nature of the solutions, since by construction $\boldsymbol{\mu} \in \boldsymbol{M}_h$ satisfies $\boldsymbol{\mu} = \mathbf{n} \times (\boldsymbol{\mu} \times \mathbf{n}) = \mu_1 \mathbf{t}_1 + \mu_2 \mathbf{t}_2$. The tangent vectors can be defined in terms of $\mathbf{n} = (n_1, n_2, n_3)$ as $\mathbf{t}_1 = (-n_2/n_1, 1, 0)$ and $\mathbf{t}_2 = (-n_3/n_1, 0, 1)$. This definition assumes that $|n_1| \geq \max(|n_2|, |n_3|)$ but analogous expressions can be obtained when $|n_2| \geq \max(|n_1|, |n_3|)$ or $|n_3| \geq \max(|n_1|, |n_2|)$ to avoid division by a small number. Boundary conditions are included by setting $\boldsymbol{M}_h(\mathbf{u}_\partial) = \{\boldsymbol{\mu} \in \boldsymbol{M}_h : \mathbf{n} \times \boldsymbol{\mu} = \Pi\mathbf{u}_\partial \text{ on } \partial\mathcal{D}\}$ and $M_h(u_\partial) = \{\mu \in M_h : \mu = \Pi u_\partial \text{ on } \partial\mathcal{D}\}$, where $\Pi\mathbf{u}_\partial$ (respectively, Πu_∂) is the projection of $\mathbf{u}_\partial$ onto $\boldsymbol{M}_h$ (respectively, u_∂ onto M_h).

Finally, we define the volume Hermitian inner products as,

$$(\eta, \zeta)_{\mathcal{T}_h} := \sum_{T \in \mathcal{T}_h} (\eta, \zeta)_T, \qquad (\boldsymbol{\eta}, \boldsymbol{\zeta})_{\mathcal{T}_h} := \sum_{i=1}^{3} (\eta_i, \zeta_i)_{\mathcal{T}_h},$$

and the surface Hermitian inner products as,

$$\langle\eta, \zeta\rangle_{\partial\mathcal{T}_h} := \sum_{T \in \mathcal{T}_h} \langle\eta, \zeta\rangle_{\partial T}, \qquad \langle\boldsymbol{\eta}, \boldsymbol{\zeta}\rangle_{\partial\mathcal{T}_h} := \sum_{i=1}^{3} \langle\boldsymbol{\eta}_i, \boldsymbol{\zeta}_i\rangle_{\partial\mathcal{T}_h}.$$

2.2. Weak formulation

In order to solve (2) on the reference domain $\Omega_\mathbf{r} \cup \overline{\Omega}_\mathbf{r}$, we first seek an approximation $(\mathbf{v}_h, \mathbf{e}_h) \in \boldsymbol{W}_h \times \boldsymbol{W}_h$ to the transformed magnetic and electric fields $(\mathbf{v}, \mathbf{e})$. Furthermore, we also introduce a new variable $\widehat{\mathbf{e}}_h$ that approximates the tangential component of the transformed electric field at the element interfaces, that is $\mathbf{n}_\mathbf{r} \times (\mathbf{e}_h \times \mathbf{n}_\mathbf{r})$. We finally introduce the approximation to the tangential component of the transformed magnetic field $\widehat{\mathbf{v}}_h$, and conservation is enforced by imposing continuity on the tangential component of $\boldsymbol{G}\widehat{\mathbf{v}}_h$ across inter-element boundaries, that is $[\![\mathbf{n}_\mathbf{r} \times \boldsymbol{G}\widehat{\mathbf{v}}_h]\!] = 0$ on $\mathcal{E}_h^o$.

The HDG method for the discretization of system (2) seeks a solution $(\mathbf{v}_h, \mathbf{e}_h, \widehat{\mathbf{e}}_h) \in \boldsymbol{W}_h \times \boldsymbol{W}_h \times \boldsymbol{M}_h(\mathbf{0})$, such that the following system holds for all $(\boldsymbol{\kappa}, \boldsymbol{\xi}, \boldsymbol{\mu}) \in \boldsymbol{W}_h \times \boldsymbol{W}_h \times \boldsymbol{M}_h(\mathbf{0})$

$$(\mathbf{v}_h, \boldsymbol{\kappa})_{\mathcal{T}_h} - (\boldsymbol{G}\mathbf{e}_h, \nabla \times \boldsymbol{\kappa})_{\mathcal{T}_h} - \langle \boldsymbol{G}\widehat{\mathbf{e}}_h, \boldsymbol{\kappa} \times \mathbf{n}_\mathbf{r} \rangle_{\partial \mathcal{T}_h} = 0,$$

$$(\boldsymbol{G}\mathbf{v}_h, \nabla \times \boldsymbol{\xi})_{\mathcal{T}_h} + \langle \boldsymbol{G}\widehat{\mathbf{v}}_h, \boldsymbol{\xi} \times \mathbf{n}_\mathbf{r} \rangle_{\mathcal{T}_h} - \omega^2 (\varepsilon \mathbf{e}_h, \boldsymbol{\xi})_{\mathcal{T}_h} = 0,$$

$$-\langle \mathbf{n}_\mathbf{r} \times \boldsymbol{G}\widehat{\mathbf{v}}_h, \boldsymbol{\mu} \rangle_{\partial \mathcal{T}_h \backslash \partial \Omega_{\mathbf{r},E}} + \langle \boldsymbol{G}\widehat{\mathbf{e}}_h, \boldsymbol{\mu} \rangle_{\partial \Omega_{\mathbf{r},E}} - i\omega \langle \sqrt{\varepsilon} \boldsymbol{G}\widehat{\mathbf{e}}_h, \boldsymbol{\mu} \rangle_{\partial \Omega_\text{rad}} = \langle \mathbf{f}, \boldsymbol{\mu} \rangle_{\partial \Omega_\mathbf{r}},$$

where

$$\mathbf{f} = \boldsymbol{G}\mathbf{e}_\partial \lceil_{\partial \Omega_{\mathbf{r},E}} - \mathbf{n} \times \boldsymbol{G}\mathbf{v}_\partial \lceil_{\partial \Omega_{\mathbf{r},V}} - \mathbf{f}_0 \lceil_{\partial \Omega_{\mathbf{r},\text{rad}}}, \tag{3}$$

and

$$\boldsymbol{G}\widehat{\mathbf{v}}_h = \boldsymbol{G}\mathbf{v}_h + \tau_t \boldsymbol{G} (\mathbf{e}_h - \widehat{\mathbf{e}}_h) \times \mathbf{n}_\mathbf{r}.$$

The global stabilization parameter τ_t is required to ensure the accuracy and stability of the HDG discretization, and defined as $\tau_t = \sqrt{\varepsilon}\omega$ [35]. Introducing the flux definition in the above system, we arrive to the final HDG system for the frequency-domain Maxwell's equations on a reference domain:

$$(\mathbf{v}_h, \boldsymbol{\kappa})_{\mathcal{T}_h} - (\boldsymbol{G}\mathbf{e}_h, \nabla \times \boldsymbol{\kappa})_{\mathcal{T}_h} - \langle \boldsymbol{G}\widehat{\mathbf{e}}_h, \boldsymbol{\kappa} \times \mathbf{n}_\mathbf{r} \rangle_{\partial \mathcal{T}_h} = 0,$$

$$(\boldsymbol{G}\mathbf{v}_h, \nabla \times \boldsymbol{\xi})_{\mathcal{T}_h} + \langle \boldsymbol{G}\mathbf{v}_h, \boldsymbol{\xi} \times \mathbf{n}_\mathbf{r} \rangle_{\mathcal{T}_h} - \omega^2 (\varepsilon \mathbf{e}_h, \boldsymbol{\xi})_{\mathcal{T}_h}$$
$$+ \langle \tau_t \boldsymbol{G}(\mathbf{e}_h - \widehat{\mathbf{e}}_h) \times \mathbf{n}_\mathbf{r}, \boldsymbol{\xi} \times \mathbf{n}_\mathbf{r} \rangle_{\mathcal{T}_h} = 0, \tag{4}$$

$$-\langle \mathbf{n}_\mathbf{r} \times \boldsymbol{G}\mathbf{v}_h + \tau_t \boldsymbol{G}\mathbf{e}_h, \boldsymbol{\mu} \rangle_{\partial \mathcal{T}_h \backslash \partial \Omega_{\mathbf{r},E}} + \langle \widetilde{\tau}_t \boldsymbol{G}\widehat{\mathbf{e}}_h, \boldsymbol{\mu} \rangle_{\partial \mathcal{T}_h} = \langle \mathbf{f}, \boldsymbol{\mu} \rangle_{\partial \Omega_\mathbf{r}},$$

where the stabilization constant $\widetilde{\tau}_t$ is given by

$$\widetilde{\tau}_t = \begin{cases} \tau_t, & \text{on } \partial \mathcal{T}_h \backslash \partial \Omega_{\mathbf{r},E} \cup \partial \Omega_{\mathbf{r},\text{rad}} \\ \tau_t - i\omega\sqrt{\varepsilon}, & \text{on } \partial \Omega_{\mathbf{r},\text{rad}} \\ 1, & \text{on } \partial \Omega_{\mathbf{r},E}. \end{cases}$$

Solving the weak formulation above is equivalent to solving the HDG discretization of (1) on the deformed domain $\mathfrak{G}(\overline{\Omega}_\mathbf{r} \cup \Omega_\mathbf{r}, \theta)$. However, we prefer the formulation in (4) since it allows us to work on a fixed parameter-independent domain and casts a clear strategy for model order reduction, as we shall describe below.

2.3. Implementation

We rewrite the weak formulation of system (4) as: find $(\mathbf{v}_h, \mathbf{e}_h, \widehat{\mathbf{e}}_h) \in \boldsymbol{W}_h \times \boldsymbol{W}_h \times \boldsymbol{M}_h(\mathbf{0})$, such that

$$\mathcal{A}(\mathbf{v}_h, \boldsymbol{\kappa}) - \mathcal{B}(\boldsymbol{G}\mathbf{e}_h, \boldsymbol{\kappa}) - \mathcal{C}(\boldsymbol{G}\widehat{\mathbf{e}}_h, \boldsymbol{\kappa}) = 0,$$

$$\mathcal{B}(\boldsymbol{G}\mathbf{v}_h, \boldsymbol{\xi}) + \mathcal{K}(\boldsymbol{G}\mathbf{v}_h, \boldsymbol{\xi}) + \mathcal{D}(\boldsymbol{G}\mathbf{e}_h, \boldsymbol{\xi}) - \omega^2 \mathcal{A}_\varepsilon(\mathbf{e}_h, \boldsymbol{\xi}) - \mathcal{E}(\boldsymbol{G}\widehat{\mathbf{e}}_h, \boldsymbol{\xi}) = 0, \tag{5}$$

$$-\mathcal{R}(\boldsymbol{G}\mathbf{v}_h, \boldsymbol{\mu}) - \mathcal{L}(\boldsymbol{G}\mathbf{e}_h, \boldsymbol{\mu}) + \mathcal{M}(\boldsymbol{G}\widehat{\mathbf{e}}_h, \boldsymbol{\mu}) = \mathcal{F}(\boldsymbol{\mu}),$$

holds for all $(\boldsymbol{\kappa}, \boldsymbol{\xi}, \boldsymbol{\mu}) \in \boldsymbol{W}_h \times \boldsymbol{W}_h \times \boldsymbol{M}_h(\mathbf{0})$. Here, the linear and bilinear forms are given by

$$\mathcal{A}(\mathbf{v}, \boldsymbol{\kappa}) = (\mathbf{v}, \boldsymbol{\kappa})_{\mathcal{T}_h}, \qquad\qquad \mathcal{A}_\varepsilon(\mathbf{e}, \boldsymbol{\xi}) = (\varepsilon \mathbf{e}, \boldsymbol{\xi})_{\mathcal{T}_h},$$

$$\mathcal{B}(\boldsymbol{G}\mathbf{e}, \boldsymbol{\kappa}) = (\boldsymbol{G}\mathbf{e}, \nabla \times \boldsymbol{\kappa})_{\mathcal{T}_h}, \qquad\qquad \mathcal{C}(\boldsymbol{G}\widehat{\mathbf{e}}, \boldsymbol{\kappa}) = \langle \boldsymbol{G}\widehat{\mathbf{e}}, \boldsymbol{\kappa} \times \mathbf{n} \rangle_{\partial \mathcal{T}_h},$$

$$\mathcal{K}(\boldsymbol{G}\mathbf{v}, \boldsymbol{\xi}) = \langle \boldsymbol{G}\mathbf{v}, \boldsymbol{\xi} \times \mathbf{n} \rangle_{\partial \mathcal{T}_h}, \qquad\qquad \mathcal{D}(\boldsymbol{G}\mathbf{e}, \boldsymbol{\xi}) = \langle \tau_t \boldsymbol{G}\mathbf{e}, \mathbf{n} \times \boldsymbol{\xi} \times \mathbf{n} \rangle_{\partial \mathcal{T}_h},$$

$$\mathcal{E}(\boldsymbol{G}\widehat{\mathbf{e}}, \boldsymbol{\xi}) = \langle \tau_t \boldsymbol{G}\widehat{\mathbf{e}}, \mathbf{n}_\mathbf{r} \times \boldsymbol{\xi} \times \mathbf{n}_\mathbf{r} \rangle_{\partial \mathcal{T}_h}, \quad \mathcal{R}(\boldsymbol{G}\mathbf{v}, \boldsymbol{\mu}) = \langle \boldsymbol{G}\mathbf{v}, \boldsymbol{\mu} \times \mathbf{n}_\mathbf{r} \rangle_{\partial \mathcal{T}_h},$$

$$\mathcal{L}(\boldsymbol{G}\mathbf{e}, \boldsymbol{\mu}) = \langle \tau_t \boldsymbol{G}\mathbf{e}, \boldsymbol{\mu} \rangle_{\partial \mathcal{T}_h}, \qquad\qquad \mathcal{M}(\boldsymbol{G}\widehat{\mathbf{e}}, \boldsymbol{\mu}) = \langle \widetilde{\tau}_t \boldsymbol{G}\widehat{\mathbf{e}}, \boldsymbol{\mu} \rangle_{\partial \mathcal{T}_h},$$

$$\mathcal{F}(\boldsymbol{\mu}) = \mathcal{F}_E(\boldsymbol{\mu}) + \mathcal{F}_V(\boldsymbol{\mu}) + \mathcal{F}_\text{rad}(\boldsymbol{\mu})$$

where the boundary form $\mathcal{F}$ arises from (3).

The discretization of the bilinear forms in (5) gives rise to the system of equations

$$
\begin{bmatrix}
\mathbb{A} & -\mathbb{B} & -\mathbb{C} \\
\mathbb{B}+\mathbb{K} & \mathbb{D}-\omega^2 \mathbb{A}_\varepsilon & -\mathbb{E} \\
-\mathbb{R} & -\mathbb{L} & \mathbb{M}
\end{bmatrix}
\begin{bmatrix}
\underline{\mathbf{v}} \\
\underline{\mathbf{e}} \\
\underline{\widehat{\mathbf{e}}}
\end{bmatrix}
=
\begin{bmatrix}
0 \\
0 \\
\mathbf{F}
\end{bmatrix}
\tag{6}
$$

for the unknowns $\underline{\mathbf{v}}, \underline{\mathbf{e}}, \underline{\widehat{\mathbf{e}}}$ corresponding to the degrees of freedom of the discretization $\mathcal{T}_h$. In practice, however, the system (6) is never assembled. Instead, we invoke the discontinuity of the approximation spaces and eliminate at the element level $\underline{\mathbf{v}}$ and $\underline{\mathbf{e}}$, also known as local variables, to obtain a system involving only the global variables $\underline{\widehat{\mathbf{e}}}$, hence reducing the size of the resulting linear system. The local variables may be rewritten as a function of the global variables using a Schur complement decomposition, namely

$$
\begin{bmatrix}
\underline{\mathbf{v}} \\
\underline{\mathbf{e}}
\end{bmatrix}
=
\begin{bmatrix}
\mathbb{A} & -\mathbb{B} \\
\mathbb{B}+\mathbb{K} & \mathbb{D}-\omega^2 \mathbb{A}_\varepsilon
\end{bmatrix}^{-1}
\begin{bmatrix}
\mathbb{C} \\
\mathbb{E}
\end{bmatrix}
\underline{\widehat{\mathbf{e}}} = \mathbb{Z}\underline{\widehat{\mathbf{e}}},
\tag{7}
$$

where $\mathbb{Z}$ has a sparse structure. That is, for a given element, the only nonzero columns correspond to the degrees of freedom of the faces of that element. The local linear systems (7) are solved during the assembly process, and $\mathbb{Z}$ is stored. Finally, the system involving only the global unknowns is given by

$$
\left(\mathbb{M} + \begin{bmatrix}\mathbb{R} & \mathbb{L}\end{bmatrix}\mathbb{Z}\right)\underline{\widehat{\mathbf{e}}} = \mathbf{F}.
\tag{8}
$$

This procedure characterizes the solution to (2) in terms of $\widehat{\mathbf{e}}_h$, which consists of only two components defined on the element faces. As a consequence, the HDG method exhibits less globally coupled degrees of freedom than other DG methods. Once the global variables $\widehat{\mathbf{e}}_h$ have been computed, the local variables $\mathbf{v}_h, \mathbf{e}_h$ can be inexpensively recovered in parallel using the pre-computed matrices $\mathbb{Z}$ for each element independently. Finally, the approximate electromagnetic fields may be recovered from the transformed fields using the identities

$$
\mathbf{V}_h = g^{-1}\mathcal{G}\mathbf{v}_h, \qquad \mathbf{E}_h = g^{-1}\mathcal{G}\mathbf{e}_h.
\tag{9}
$$

3. Reduced order modeling

The formulation and implementation introduced above describe the procedure for computing solutions to (2) for any combination of the parameter values $(\omega, \varepsilon, \theta)$. In order to describe the construction of ROM, we first rewrite the HDG system (6) in a more compact form as follows: find a solution $\mathbf{e}_h^\dagger := (\mathbf{v}_h, \mathbf{e}_h, \widehat{\mathbf{e}}_h)$ in the $\mathcal{N}^\dagger$-dimensional approximation space $\boldsymbol{W}_h^\dagger := \boldsymbol{W}_h \times \boldsymbol{W}_h \times \boldsymbol{M}_h(\mathbf{0})$ satisfying

$$
\mathcal{A}^\dagger\left(\mathbf{e}_h^\dagger, \boldsymbol{\xi}^\dagger; (\omega, \varepsilon, \theta)\right)\mathbf{e}_h^\dagger = \mathcal{F}^\dagger\left(\boldsymbol{\xi}^\dagger; (\omega, \varepsilon, \theta)\right),
\tag{10}
$$

for any $\boldsymbol{\xi}^\dagger := (\boldsymbol{\kappa}, \boldsymbol{\xi}, \boldsymbol{\mu}) \in \boldsymbol{W}_h^\dagger$.

Let us now assume that we are given a collection of solutions $\mathbf{e}_h^\dagger \in \boldsymbol{W}_h^\dagger$, or snapshots, computed at multiple parameter values through expressions (7) and (8). In order to extract a low-dimensional representation of the parametric dependence, we apply POD and obtain a set of $N_{\max}$ orthonormalized basis functions $\boldsymbol{\zeta}_n^\dagger \in \boldsymbol{W}_h^\dagger$, $1 \le n \le N_{\max}$ such that $(\boldsymbol{\zeta}_n^\dagger, \boldsymbol{\zeta}_{n'}^\dagger)_{\boldsymbol{W}^\dagger} = \delta_{nn'}$ for $1 \le n, n' \le N_{\max}$, where the inner product $(\cdot, \cdot)_{\boldsymbol{W}^\dagger}$ is defined as

$$
(\boldsymbol{\xi}_1^\dagger, \boldsymbol{\xi}_2^\dagger)_{\boldsymbol{W}^\dagger} = (\boldsymbol{\kappa}_1, \boldsymbol{\kappa}_2)_{\mathcal{T}_h} + (\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)_{\mathcal{T}_h} + \langle\boldsymbol{\mu}_1, \boldsymbol{\mu}_2\rangle_{\partial\mathcal{T}_h}.
$$

The choice $N_{\max}$ is made by monitoring the ratio between the energy in the model to the total energy of the snapshot matrix above using singular value information, see [56] for more details.

The orthonormalized basis functions allow us to define an associated hierarchical POD space $\boldsymbol{W}_N^\dagger$ as

$$
\boldsymbol{W}_N^\dagger = \mathrm{span}\{\boldsymbol{\zeta}_n^\dagger, \ 1 \le n \le N\}, \quad N = 1, \ldots, N_{\max}.
$$

We now consider the Galerkin variational statement (10) but restricted to $\boldsymbol{W}_N^\dagger \subset \boldsymbol{W}_h^\dagger$, that is: for an unseen parametric tuple $(\omega, \varepsilon, \theta)$, find an approximate solution $\mathbf{e}_N^\dagger \in \boldsymbol{W}_N^\dagger$ satisfying

$$
\mathcal{A}^\dagger\left(\mathbf{e}_N^\dagger, \boldsymbol{\xi}^\dagger; (\omega, \varepsilon, \theta)\right)\mathbf{e}_N^\dagger = \mathcal{F}^\dagger\left(\boldsymbol{\xi}^\dagger; (\omega, \varepsilon, \theta)\right),
\tag{11}
$$

for any $\boldsymbol{\xi}^\dagger \in \boldsymbol{W}_N^\dagger$. The above system of equations is low-dimensional and, in principle, we would expect that its evaluation for a new set of parameters is inexpensive. The main challenge, however, is that the geometry parameter θ enters in the definitions of $\mathcal{A}^\dagger$ and $\mathcal{F}^\dagger$ through $\boldsymbol{G}$ in a *non-affine* manner. This prevents us from isolating the parametric dependence and precomputing the required integrals required for the evaluation of $\mathcal{A}^\dagger$ and $\mathcal{F}^\dagger$. This means that solving system (11) does not offer substantial computational advantages compared to the regular HDG system (10), since it still requires formation and assembly of the $\boldsymbol{G}$-dependent bilinear forms in (5) for any new θ. In order to circumvent this limitation, we resort to empirical interpolation techniques.
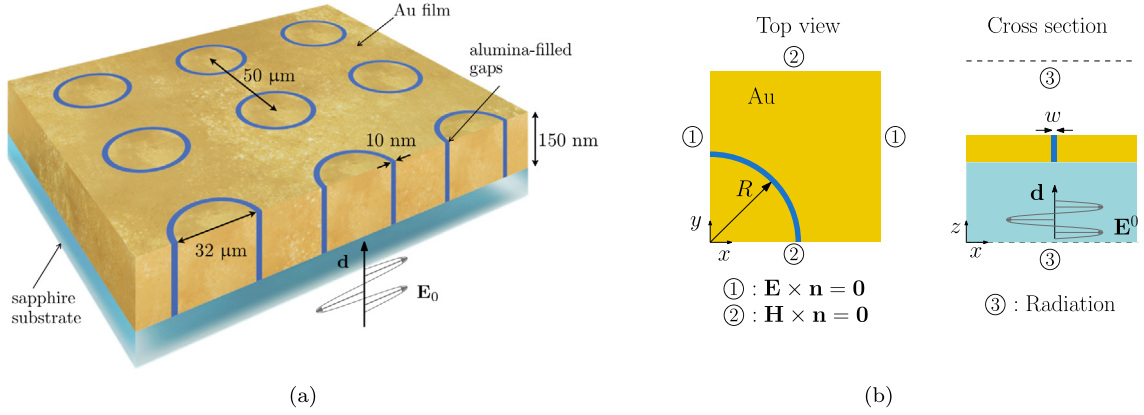
Fig. 3. (a) Schematic diagram of thin gold film on silica substrate patterned with periodic square array of alumina gaps under plane wave THz illumination. (b) Top and cross section of computational domain, with dimensions and boundary conditions.

3.1. Empirical interpolation

A widespread technique in function interpolation is EIM [2]. The idea behind EIM is to approximate a non-affine parametrized function $f(\mathbf{x}, \theta)$ by a weighted combination of $\mathcal{Q}$ orthogonal spatial functions $\mathbf{\Phi} = [\phi_1(\mathbf{x}), \ldots, \phi_{\mathcal{Q}}(\mathbf{x})]$, that is $f(\mathbf{x}, \theta) \approx \mathbf{\Phi}(\mathbf{x})\boldsymbol{\alpha}(\theta)$ for a certain parameter-dependent coefficient vector $\boldsymbol{\alpha}(\theta)$.

In order to apply the EIM to the bilinear forms (5), we first need to identify the non-affine parametrized functions to be interpolated, as well as the discrete set of spatial points $\mathbf{x}$ where the interpolants will be evaluated. In the spirit of keeping the main body of the article as clear as possible, we defer the implementation details of EIM to Appendix A, and assume from this point onwards that both $\mathcal{A}^{\dagger}$ and $\mathcal{F}^{\dagger}$ may be approximated as

$$\mathcal{A}^{\dagger}\left(\cdot, \cdot ;(\omega, \varepsilon, \theta)\right) \approx \sum_{q=1}^{\mathcal{Q}} \alpha_q^{\mathcal{A}}(\theta) \mathcal{A}_q^{\dagger}\left(\cdot, \cdot ;(\omega, \varepsilon)\right)$$

$$\mathcal{F}^{\dagger}\left(\cdot ;(\omega, \varepsilon, \theta)\right) \approx \sum_{q=1}^{\mathcal{Q}} \alpha_q^{\mathcal{F}}(\theta) \mathcal{F}_q^{\dagger}\left(\cdot ;(\omega, \varepsilon)\right)$$

(12)

where the $\mathcal{A}_q^{\dagger}$, $\mathcal{F}_q^{\dagger}$ forms are affine in ω and ε.

If we now insert the approximations (12) into (11), we arrive at a linear system that is *affine* in both the geometric parameters θ and the optical parameters ω, ε by construction, that is

$$\sum_{q=1}^{\mathcal{Q}} \alpha_q^{\mathcal{A}}(\theta) \mathcal{A}_q^{\dagger}\left(\mathbf{e}_N^{\dagger}, \boldsymbol{\xi}^{\dagger} ;(\omega, \varepsilon)\right) \mathbf{e}_N^{\dagger} = \sum_{q=1}^{\mathcal{Q}} \alpha_q^{\mathcal{F}}(\theta) \mathcal{F}_q^{\dagger}\left(\boldsymbol{\xi}^{\dagger} ;(\omega, \varepsilon)\right),$$

with solution $\mathbf{e}_N^{\dagger} \in \mathbf{W}_N^{\dagger}$ for any $\boldsymbol{\xi}^{\dagger} \in \mathbf{W}_N^{\dagger}$.

Furthermore, for each new tuple $(\omega, \varepsilon, \theta)$ the system can be assembled extremely efficiently, since the dimension of the matrices is $N \times N$ due to the Galerkin projection using the low-dimensional approximation space $W_N^{\dagger}$. The solution of this reduced order system are the trial coefficients $\{\lambda_n^{\dagger}\}_{n=1}^N$, that is $\mathbf{e}_N^{\dagger} = \sum_{n=1}^N \lambda_n^{\dagger} \zeta_n^{\dagger}$, enabling us to recover the approximate transformed electromagnetic fields as $\mathbf{e}_N^{\dagger} = (\mathbf{v}_N, \mathbf{e}_N, \widehat{\mathbf{e}}_N)$. The actual approximate electromagnetic fields $\mathbf{V}_N$, $\mathbf{E}_N$ are obtained from $\mathbf{v}_N$, $\mathbf{e}_N$ using (9) element-wise. Note that, for efficiency purposes, the non-affine function $g^{-1}\mathcal{G}$ needed for the transformation can also be approximated with the EIM. Details concerning computational cost and implementation strategy are also described in Appendix A.

4. Numerical results

4.1. 3D periodic coaxial nanogap structure

The first example consists of a 3D periodic annular structure shown in Fig. 3a, which has been used for plasmonic sensing applications as well as fundamental studies of nanophotonics phenomena [42,59]. A 150 nm-thick gold film is deposited over a sapphire substrate, and the coaxial nanoapertures are filled with Al_2O_3. For this problem, we choose the diameter of the rings to be 32 microns with square periodicity of 50 μm, and a gap width of 10 nm.

Table 1
Variability ranges, interval of validity and references for parameters of structure in Fig. 3a.

Parameter	Values	Wavelength range (μm)	References
$\hbar\overline{\omega}_p$ [eV]	9.02 ± 0.18	< 12	[39,41,48]
$\hbar\overline{\gamma}$ [eV]	0.02678 ± 0.007	< 12	[39,41,48]
$\varepsilon_{Al_2O_3}$	5–6	thickness-dependent	[21]
$n_{sapphire}$	3.07 ± 0.006	< 2	[20]

We illuminate the annular structure from below with an *x*-polarized plane wave, and for efficiency purposes we reduce the computational domain by exploiting the symmetries of the lattice, hence we only need to solve for one quadrant of the ring structure as indicated in Fig. 3b. Symmetries are imposed by setting $\mathbf{E} \times \mathbf{n} = \mathbf{0}$ on the *x*-constant boundaries and $\mathbf{H} \times \mathbf{n} = \mathbf{0}$ on the *y*-constant boundaries, whereas radiation conditions are prescribed on the *z*-constant boundaries.

To analyze the performance of the structure, we focus on the electric field's *x*-component enhancement π within the gap volume and the transmitted power ς through the structure, computed as

$$\pi = \frac{\int_{gap} |\mathbf{E}_x| \, dV}{\int_{gap} |\mathbf{E}_{0,x}| \, dV}, \qquad \varsigma = \frac{\int_{A_1} |\Re \left[\mathbf{E} \times \mathbf{H}^* \right] \cdot \mathbf{n}| \, dA}{\int_{A_0} |\Re \left[\mathbf{E}_0 \times \mathbf{H}_0^* \right] \cdot \mathbf{n}| \, dA}, \tag{13}$$

where A_0 is an arbitrary *xy* plane below the gold film and A_1 an arbitrary *xy* plane above the gold film.

The HDG discretization is composed of 1.3K hexahedral anisotropic elements, enabling us to accommodate the multiple length scales present in the structure with a reduced number of degrees of freedom, where a cubic representation of the geometry is employed to better represent the curved annular boundaries. Further details on the meshing strategies for these periodic coaxial nanostructures may be found in [55]. For the examples in this article, numerical accuracy is verified through grid convergence studies against a very fine mesh, until the relative error for the field enhancement is below 0.1% at resonance.

4.1.1. Material parameters

One of the difficulties of simulating plasmonic devices is the determination of the permittivities for metals and dielectrics, since data that comes from measurements is often noisy and may exhibit significant variability. To that end, we develop a ROM considering the material properties as exploration parameters. The main objective is to show that, with few full model evaluations, we can construct a ROM that is capable of predicting inexpensively the response of the plasmonic device for a range of material properties and frequencies. In this example, we assume there are no geometry deformations, that is $\mathfrak{G} = $ id.

Firstly, we define the parameters and their corresponding intervals of variability. We are interested in studying the field enhancement on the aperture volume of the coaxial structure introduced in Fig. 3a, with frequencies ranging from 0.5 THz to 0.9 THz. The range of values used for the relevant parameters, along with the references wherefrom data was extracted, is summarized in Table 1.

Secondly, the matrix of snapshots is obtained by sampling the 5-dimensional parameter space – ω plus four material parameters in Table 1 – and computing the full electromagnetic solution at 200 selected parameter values. The parameter values are chosen according to the Sobol sequence [52] in order to achieve a more uniform sampling in the high-dimensional space.

In order to assess the accuracy of the POD basis, we compute the HDG solution for 100 random parameter values – referred to as the test set – and evaluate the relative error committed by the reduced order model on the test set of examples using up to $N = 30$ modes in the basis. The relevant statistics of the distribution of relative errors clearly shows that the accuracy of the reduced model increases with the size of the basis N, see Fig. 4. Furthermore, with only 10 modes 95% of all testing examples are already below the threshold 0.1%, which is generally deemed sufficient for most engineering applications. Note that the ROM does not converge exponentially for this problem, which may be due to the high dimensionality of the parameter space and the non-regularity of the quantity of interest.

The efficiency of the ROM is assessed by timing its online performance, more specifically the assembly, solution of the linear system and recovery of the local variables $(\mathbf{V}_N, \mathbf{E}_N)$ from the trial coefficients $\{\lambda_n^\dagger\}_{n=1}^N$. Time estimates are obtained averaging the wall time for 500 runs for each task using a single processor of a 512GB `Linux` 12.04 machine with 32 AMD Opteron(tm) Processors 6320x15, and results are collected in Table 2. The advantages of the reduced order model are remarkable for this 3D problem, since we achieve an online cost reduction of ~ 4 orders of magnitude for the most expensive POD basis without compromising accuracy, as seen in Fig. 4. As anticipated, the bulk of the computational cost to evaluate the POD basis is devoted to recovering the local variables required to evaluate the quantity of interest, since it involves operating with the high-dimensional POD basis functions. This approach is particularly beneficial for large problems with parametric variability, where the model must produce accurate solutions in real-time for multiple queries.

The comparison above concerns only the online cost, but for completeness we shall now discuss the offline cost as well. The bulk of the offline cost is spent computing the HDG solutions that are later employed to construct the POD approximation space $\mathbf{W}_N^\dagger$. Note however, that this solutions can be easily computed in parallel. Since it is impossible to
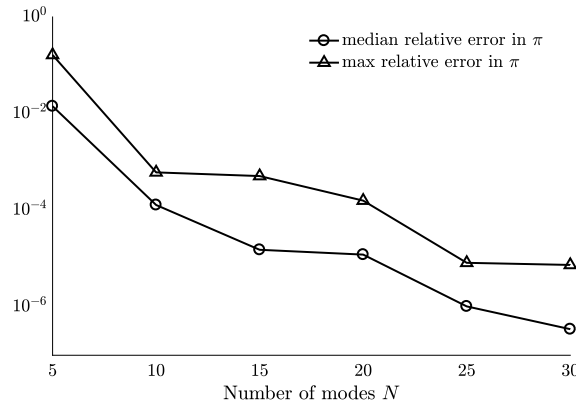
Fig. 4. Median and maximum relative error on field enhancement for the parameters in the test set, evaluated using 6 reduced models with increasing fidelity.

Table 2
Times in s·10^{-4}. Reduced model can be evaluated $\mathcal{O}(10^4)$ times faster overall than HDG while maintaining accuracy.

Model	Assembly	Linear system	$\mathbf{V}_N, \mathbf{E}_N$
$N = 5$	0.64	0.83	44.5
$N = 10$	0.73	1.13	48.6
$N = 15$	0.94	1.17	51.7
$N = 20$	1.10	1.42	56.9
$N = 25$	1.36	1.75	58.0
HDG	7.6e5	1.8e6	8e3

establish *a priori* the optimal number of snapshots required, we typically start with a small number of snapshots and then gradually enrich the solution space if the accuracy requirements on the test set are not met. All in all, creating a ROM for a 3D realistic problem is a computationally demanding task, hence it is only justified if the resulting ROM ought to be queried multiple times.

4.1.2. Geometry deformations

In this section, we demonstrate the effectivity of ROM for geometry deformations on the periodic coaxial structure introduced above. The objective is to show that, when the domain is subject to deformations, instead of solving Maxwell's equations on the deformed structure, it is desirable to solve the transformed Maxwell's equations (2) on the original reference structure in Fig. 3b.

We analyze the effect of modifying the radius and gap width of the rings, with nominal values $R = 16$ μm and $w = 10$ nm. The modifications considered in this paper, for both radius and gap, are homogeneous with respect to the angle of the ring, hence $\theta = (R, w)$. The material parameters and gold optical constants are set to $\varepsilon_{\text{Al}_2\text{O}_3} = 5.5$, $n_{\text{sapphire}} = 3.07$, $\hbar\omega_p = 9.02$ eV, $\hbar\gamma = 0.02678$ eV, $\varepsilon_\infty = 1$, and we study two different models: a $\pm30\%$ variation for the radius, that is $R \in [11.2, 20.8]$ μm, for a fixed gap $w = 10$ nm, denoted as M_1; and a $\pm20\%$ variation for both the radius and the gap, namely $R \in [12.8, 19.2]$ μm and $w \in [8, 12]$ nm, denoted as M_2. Finally, we consider frequencies ranging from 0.3 THz to 0.9 THz.

The diffeomorphism $\mathfrak{G}$ prescribes variations of both the radius and the gap. The mapping requires continuity of the derivatives for the entire domain in order to ensure the face integrals in the weak formulation are well defined. To that end, we use $\mathcal{C}^2$ cubic splines [12] in the radial direction to construct the mapping, described in detail in Appendix B.

The POD basis for M_1 (resp. M_2) is formed by computing 250 (resp. 400) solutions of (4) at different (ω, θ) values, and then compressing them to form a basis of 98 (resp. 86) modes. Once the ROM is constructed, it can be efficiently queried for any valid combination of frequency and radius or radius-gap. We collect the wall time elapsed to evaluate the different pieces of the ROM. Recovering the empirical interpolation coefficients $\boldsymbol{\alpha}(\theta)$ is independent of the dimension of the ROM. Similarly as before, the bulk of the cost is incurred calculating the approximate transformed electromagnetic fields $\mathbf{v}_N, \mathbf{e}_N$ from the trial coefficients, since it involves operating with the high-fidelity dimension $\mathcal{N}^\dagger$. Moreover, the reference domain formulation requires non negligible additional computations, since the approximate field variables $\mathbf{V}_N, \mathbf{E}_N$ need to be recovered through (9). All in all, the benefits of employing a ROM to evaluate quantities of interest of the full EM simulation are remarkable, since a speedup of more than 3 orders of magnitude is achieved, even for the largest models, despite the overhead incurred by the empirical interpolation procedure, see Table 3.

Table 3
Times in ms. Reduced model can be evaluated $\mathcal{O}(10^3)$ times faster overall than HDG despite the EIM approximation.

Model	N	EIM coefficients	Assembly	Linear system	$\mathbf{v}_N$, $\mathbf{e}_N$	$\mathbf{V}_N$, $\mathbf{E}_N$
M_1	25	21.78	1.83	0.21	2.97	4.90
	50		3.68	0.33	3.32	
	75		12.55	0.68	4.96	
	98		23.05	1.03	5.82	
M_2	25	21.85	2.07	0.18	2.60	4.85
	50		9.37	0.35	3.87	
	71		20.25	0.66	4.89	
	86		26.45	0.81	5.45	
HDG		–	1.1e5	7.2e5	9.5e2	

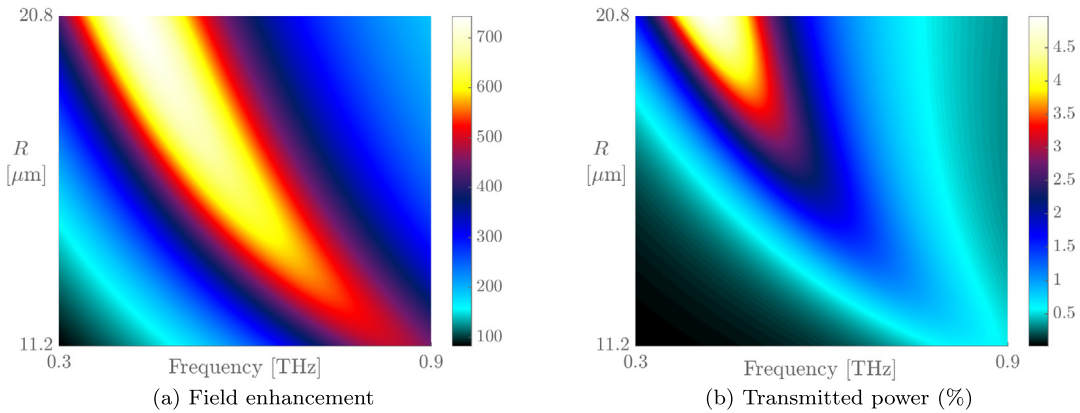(a) Field enhancement

(b) Transmitted power (%)

Fig. 5. Resonant quantities maps for frequency and radius, evaluated with the M_1 reduced order model. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

Finally, we go one step further and show how ROM can be leveraged to achieve a deeper understanding of this structure. For instance, we can analyze the field enhancement and transmission profiles as a function of both the frequency and the radius of the annular aperture. For a fixed gap size, reducing the radius has a significant impact in the resonant quantities, more specifically blue-shifting the resonance, as shown in Fig. 5.

Model M_2 gives rise to additional interpretations and results, since we can not only study the impact of both the radius and the gap width separately, but also their interactions. The value of the resonant wavelength λ^* for the nominal gap width 10 nm as a function of the radius is depicted in Fig. 6a. Additionally, the λ^* dependence on the gap size for a fixed radius is shown in Fig. 6c. For both cases, we also provide intervals that correspond to the sensitivities of λ^* with respect to $\pm 5\%$, $\pm 10\%$ and $\pm 20\%$ relative variation of the either gap δw or the radius δR. The ROM is essential to compute these results, since for each ring radius value and gap width a frequency sweep to detect the resonant quantities is required.

Furthermore, additional relevant information may be extracted from the ROM. One such example is the relationship that maps gap relative variations δw (resp. radius relative variations δR) to relative shifts in resonant wavelength $\delta \lambda^*$, see Fig. 6b (resp. Fig. 6d). These mappings are obtained effortlessly once the ROM is available, and yet contribute to acquiring a more profound understanding of the device's behavior.

4.2. Design of a concentric annular structure

For this example, we will employ the ROM framework to investigate the annular structure with two concentric rings under several design criteria, and compare its performance with a single ring structure at low THz frequencies. Structures consisting of concentric coaxial nanogaps are expected to exhibit superior performance, since resonances are excited at distinct frequencies. More interestingly, the question of how to arrange the concentric apertures arises naturally, as different configurations may lead to significant changes in transmission.

The reference structure is shown in Fig. 7, with ring diameters of 26 and 38 microns. The material parameters and gold optical constants are set to $\varepsilon_{Al_2O_3} = 5.5$, $n_{sapphire} = 3.07$, $\hbar \omega_p = 9.02$ eV, $\hbar \gamma = 0.02678$ eV, $\varepsilon_\infty = 1$, and we consider frequencies in the range 0.3 THz to 0.9 THz. The gap widths are fixed to 10 nm, thus the only tunable geometric parameters are the radii of the rings $\theta = (R_1, R_2)$, set as $\pm 10\%$ of the nominal value, that is $R_1 \in [11.7,\ 14.3]$ μm and $R_2 \in [17.1,\ 20.9]$ μm. We then construct a surrogate of the high-fidelity model, using 650 snapshots computed over the three-dimensional parametric space formed by the ω and θ, and results in a ROM with 74 modes.
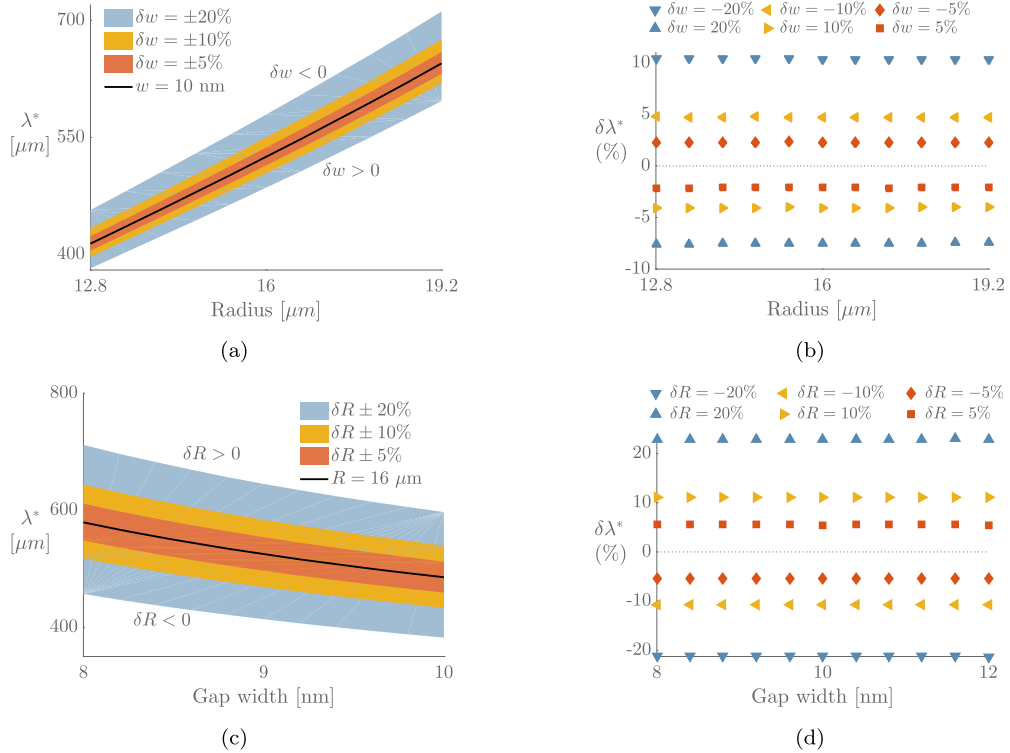
Fig. 6. Absolute and relative sensitivities of resonant quantities on geometry modifications with M_2. (a)–(b) Gap variations and resonant wavelength. (c)–(d) Radius variations and resonant wavelength.
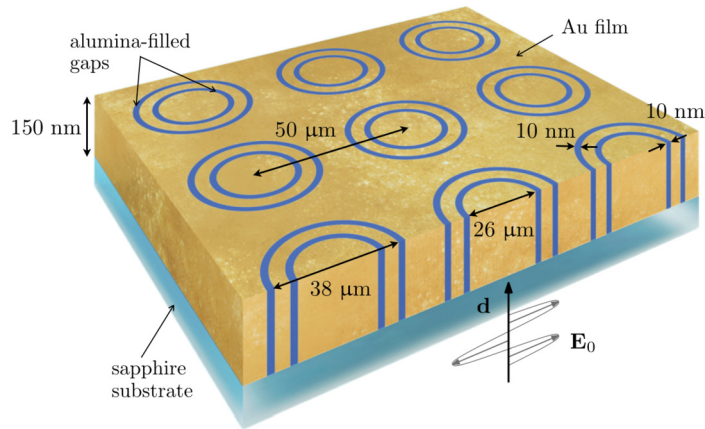

Fig. 7. Schematic diagram of thin gold film on sapphire substrate patterned with periodic square array of concentric alumina gaps under plane wave THz illumination.

In order to illustrate the potential of the concentric coaxial structure, we shall investigate three distinct objective functions to drive the design process, namely

$$\mathcal{H}_1 = \max_{R_1,R_2} \varsigma, \qquad \mathcal{H}_2 = \max_{R_1,R_2} E_f[\varsigma] - \sqrt{V_f[\varsigma]}, \qquad \mathcal{H}_3 = \max_{R_1,R_2} E_f[\varsigma] - 10\sqrt{V_f[\varsigma]}$$

where E_f, V_f refer to the expectation and variance respectively, for frequencies f in the range 0.4 THz to 0.7 THz. The objective function $\mathcal{H}_1$ seeks to maximize the transmitted power along the spectrum considered, whereas both $\mathcal{H}_2$, $\mathcal{H}_3$ target radii configurations that are robust to frequency variations (with distinct penalties) within the interval of interest. Fortunately, since the design space is only bidimensional it suffices to discretize it using a fine step size and compute the objective function invoking the ROM for all possible combinations of radii, thus avoiding the use of optimization algorithms.

The objective functions normalized between 0 and 1 along with the maximizing parameter configuration (marked with a cross) are represented in Fig. 8a. For each objective function, the transmitted power ς spectrum for the optima radii
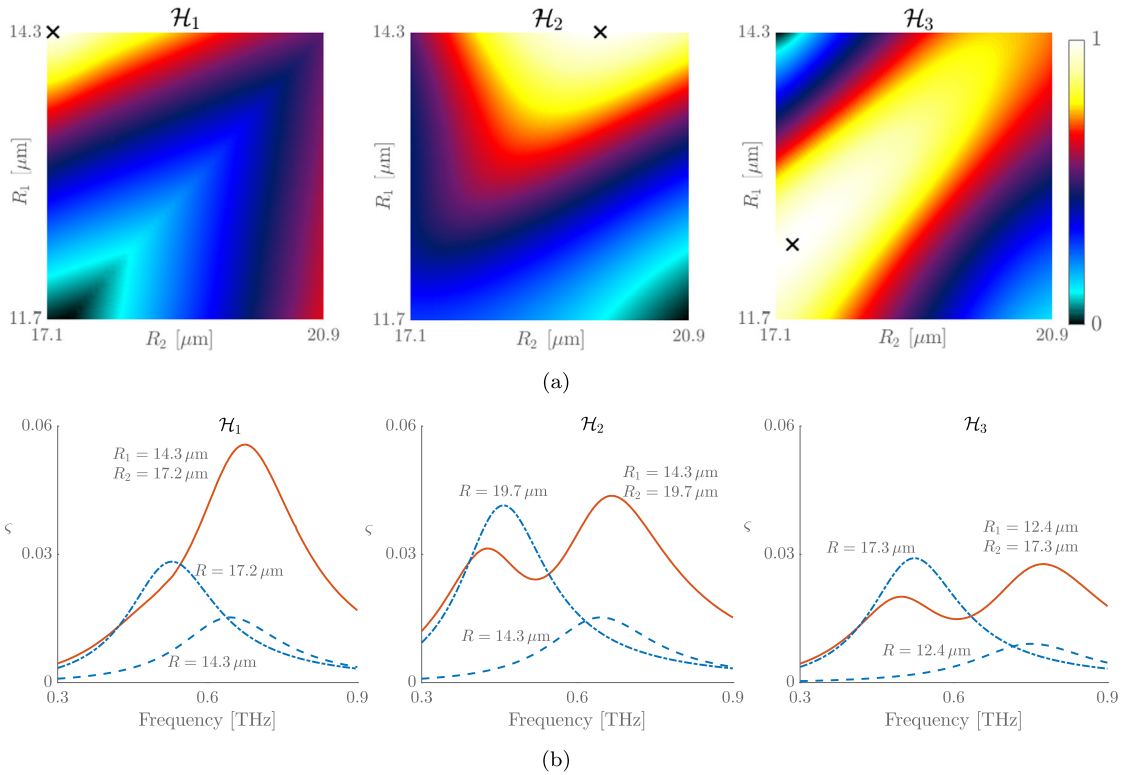
(a)

(b)

Fig. 8. (a) Objective functions shown as surface plot for inner R_1 and outer R_2 radii values, normalized to [0, 1]. Optimal configuration is shown with black cross. (b) Frequency-transmission profiles of optimal configurations for different objective functions.

(solid line), as well as the power transmitted by a structure with a single ring, for both the inner (dashed) and the outer (dash-dot) radius, are collected in Fig. 8b. Note that the response of the concentric ring structure differs significantly from the responses of the single ring configurations. Maximum transmission is attained when the concentric rings are separated by a small distance, whereby the transmission peaks that would correspond to each ring fuse into a single enhanced resonance. Moreover, the optimal configuration shown in Fig. 8a (left) suggests that larger transmissions may be encountered if the separation is further reduced.

Conversely, if robust optima are sought, the ideal radii configurations result in increased separation, enabling the clear identification of the resonances corresponding to both annular apertures. The main changes between $\mathcal{H}_2$ and $\mathcal{H}_3$ is that the former prioritizes a larger average transmission on the interval of interest, whereas the latter renders a design less sensitive to frequency changes, at the expense of smaller average transmission. In both cases, the concentric ring structure provides increased transmission over a broader range of frequencies than that of the single ring.

5. Conclusions

In this paper, we have developed a reduced order modeling framework for plasmonic coaxial nanogap structures based on the HDG method for Maxwell's equations. In order to account for the variability in the geometry of the plasmonic structures, we have proposed an alternative formulation of the HDG method for Maxwell's equations on a reference (parameter-independent) domain. The reference formulation, combined with proper orthogonal decomposition and empirical interpolation techniques, enables us to incorporate both material and geometric variability into the reduced order model. Furthermore, the ROM offline-online decomposition strategy allows us to compute solutions to Maxwell's equations inexpensively. The developed tool is advantageous for real design situations where multiple solutions for different values of the design parameters are required.

Acknowledgements

Appendix A. EIM on non-affine functions of HDG discretization

In this appendix, we provide extensive documentation on how to perform discrete empirical interpolation on the bilinear forms introduced in (5), thus enabling the approximation of the linear system in (11) using the affine expression (12). In addition, we review the computational strategy pursued to efficiently evaluate the reduced order model.

Discretization of bilinear forms

Firstly, for the discretization of the forms in (5), we introduce φ_i, $1 \leq i \leq I$ to be the basis functions of $\mathcal{P}^p(T)$ and ϕ_k, $1 \leq k \leq K$ the basis functions of $\mathcal{P}^p(F)$. Bearing in mind that the 3x3 deformation tensor $\boldsymbol{G}$ is symmetric, we derive the contribution from an arbitrary element $T \in \mathcal{T}_h$ to the global system. Furthermore, we shall make use of the identity $\boldsymbol{\mu} = \mathbf{n_r} \times (\boldsymbol{\mu} \times \mathbf{n_r})$, for $\boldsymbol{\mu} \in \boldsymbol{M}_h(\mathbf{0})$ spanned by the reference-domain tangent vectors $\mathbf{t}_1$ and $\mathbf{t}_2$.

The elemental matrices for element T have the following form:

$$
\mathbb{A}^t = \begin{bmatrix} \mathbb{A}_0^t & 0 & 0 \\ 0 & \mathbb{A}_0^t & 0 \\ 0 & 0 & \mathbb{A}_0^t \end{bmatrix} \qquad
\mathbb{B}^t = \begin{bmatrix} \mathbb{B}_{11}^t & \mathbb{B}_{12}^t & \mathbb{B}_{13}^t \\ \mathbb{B}_{21}^t & \mathbb{B}_{22}^t & \mathbb{B}_{23}^t \\ \mathbb{B}_{31}^t & \mathbb{B}_{32}^t & \mathbb{B}_{33}^t \end{bmatrix} \qquad
\mathbb{K}^t = \begin{bmatrix} \mathbb{K}_{11}^t & \mathbb{K}_{12}^t & \mathbb{K}_{13}^t \\ \mathbb{K}_{21}^t & \mathbb{K}_{22}^t & \mathbb{K}_{23}^t \\ \mathbb{K}_{31}^t & \mathbb{K}_{32}^t & \mathbb{K}_{33}^t \end{bmatrix}
$$

$$
\mathbb{C}^t = \begin{bmatrix} \mathbb{C}_{11}^t & \mathbb{C}_{12}^t \\ \mathbb{C}_{21}^t & \mathbb{C}_{22}^t \\ \mathbb{C}_{31}^t & \mathbb{C}_{32}^t \end{bmatrix} \qquad
\mathbb{D}^t = \begin{bmatrix} \mathbb{D}_{11}^t & \mathbb{D}_{12}^t & \mathbb{D}_{13}^t \\ \mathbb{D}_{21}^t & \mathbb{D}_{22}^t & \mathbb{D}_{23}^t \\ \mathbb{D}_{31}^t & \mathbb{D}_{32}^t & \mathbb{D}_{33}^t \end{bmatrix} \qquad
\mathbb{E}^t = \begin{bmatrix} \mathbb{E}_{11}^t & \mathbb{E}_{12}^t \\ \mathbb{E}_{21}^t & \mathbb{E}_{22}^t \\ \mathbb{E}_{31}^t & \mathbb{E}_{32}^t \end{bmatrix} \qquad \text{(A.1)}
$$

$$
\mathbb{R}^t = \begin{bmatrix} \mathbb{R}_{11}^t & \mathbb{R}_{12}^t & \mathbb{R}_{13}^t \\ \mathbb{R}_{21}^t & \mathbb{R}_{22}^t & \mathbb{R}_{23}^t \end{bmatrix} \qquad
\mathbb{L}^t = \begin{bmatrix} \mathbb{L}_{11}^t & \mathbb{L}_{12}^t & \mathbb{L}_{13}^t \\ \mathbb{L}_{21}^t & \mathbb{L}_{22}^t & \mathbb{L}_{23}^t \end{bmatrix} \qquad
\mathbb{M}^t = \begin{bmatrix} \mathbb{M}_{11}^t & \mathbb{M}_{12}^t \\ \mathbb{M}_{21}^t & \mathbb{M}_{22}^t \end{bmatrix}
$$

Below, we provide expressions for the different subblocks for each elemental matrix, using for the volume φ (resp. face ϕ) basis functions i/j (resp. k/ℓ) as test/trial indices respectively, $1 \leq c, d \leq 3$ for the dimension indices and $1 \leq a, b \leq 2$ for the tangent vectors indices. The matrix $\mathbb{A}^t$ consists of mass matrix subblocks $\mathbb{A}_{0,ij}^t = (\varphi_i, \varphi_j)_T$ in the diagonal. The different components of the curl-convection matrices in $\mathbb{B}^t$ are obtained as

$$
\begin{aligned}
\mathbb{B}_{1d,ij}^t &= (\partial_z \varphi_i, \boldsymbol{G}_{2d}\, \varphi_j)_T - (\partial_y \varphi_i, \boldsymbol{G}_{3D}\, \varphi_j)_T\,, \\
\mathbb{B}_{2d,ij}^t &= (\partial_x \varphi_i, \boldsymbol{G}_{3D}\, \varphi_j)_T - (\partial_z \varphi_i, \boldsymbol{G}_{1d}\, \varphi_j)_T\,, \\
\mathbb{B}_{3D,ij}^t &= (\partial_y \varphi_i, \boldsymbol{G}_{1d}\, \varphi_j)_T - (\partial_x \varphi_i, \boldsymbol{G}_{2d}\, \varphi_j)_T\,.
\end{aligned} \qquad \text{(A.2)}
$$

The subblocks for $\mathbb{K}^t$ and $\mathbb{D}^t$ are $\mathbb{K}_{cd,ij}^t = \langle \kappa_{cd}\varphi_i, \varphi_j \rangle_{\partial T}$ and $\mathbb{D}_{cd,ij}^t = \langle \delta_{cd}\varphi_i, \varphi_j \rangle_{\partial T}$, where κ_{cd}, δ_{cd} are given by

$$
\kappa_{1d} = \boldsymbol{G}_{3D}n_2 - \boldsymbol{G}_{2d}n_3\,, \quad \kappa_{2d} = \boldsymbol{G}_{1d}n_3 - \boldsymbol{G}_{3D}n_1\,, \quad \kappa_{3D} = \boldsymbol{G}_{2d}n_1 - \boldsymbol{G}_{1d}n_2\,, \qquad \text{(A.3)}
$$

$$
\begin{aligned}
\delta_{1d} &= \boldsymbol{G}_{1d}(1 - n_1^2) - n_1 n_3 \boldsymbol{G}_{3D} - n_1 n_2 \boldsymbol{G}_{2d}\,, \\
\delta_{2d} &= \boldsymbol{G}_{2d}(1 - n_2^2) - n_1 n_2 \boldsymbol{G}_{1d} - n_2 n_3 \boldsymbol{G}_{3D}\,, \\
\delta_{3D} &= \boldsymbol{G}_{3D}(1 - n_3^2) - n_1 n_3 \boldsymbol{G}_{1d} - n_2 n_3 \boldsymbol{G}_{2d}\,.
\end{aligned} \qquad \text{(A.4)}
$$

The submatrices for $\mathbb{C}^t$, $\mathbb{E}^t$ are given by $\mathbb{C}_{cb,i\ell}^t = \langle \gamma_{cb}\varphi_i, \phi_\ell \rangle_{\partial T}$, $\mathbb{E}_{cb,i\ell}^t = \langle \epsilon_{cb}\varphi_i, \phi_\ell \rangle_{\partial T}$. For simplicity, we introduce the modified tangent vectors $\widehat{\mathbf{t}}_b = \boldsymbol{G}\mathbf{t}_b$, and define γ_{cb}, ϵ_{cb} as

$$
\gamma_{1b} = \widehat{t}_{b3}n_2 - \widehat{t}_{b2}n_3\,, \quad \gamma_{2b} = \widehat{t}_{b1}n_3 - \widehat{t}_{b3}n_1\,, \quad \gamma_{3b} = \widehat{t}_{b2}n_1 - \widehat{t}_{b1}n_2\,, \qquad \text{(A.5)}
$$

$$
\begin{aligned}
\epsilon_{1b} &= \widehat{t}_{b1}(1 - n_1^2) - n_1 n_3 \widehat{t}_{b3} - n_1 n_2 \widehat{t}_{b2}\,, \\
\epsilon_{2b} &= \widehat{t}_{b2}(1 - n_2^2) - n_1 n_2 \widehat{t}_{b1} - n_2 n_3 \widehat{t}_{b3}\,, \\
\epsilon_{3b} &= \widehat{t}_{b3}(1 - n_3^2) - n_1 n_3 \widehat{t}_{b1} - n_2 n_3 \widehat{t}_{b2}\,.
\end{aligned} \qquad \text{(A.6)}
$$

In addition, the $\mathbb{R}^t$ components are given by $\mathbb{R}_{ad,kj}^t = \langle [\boldsymbol{G}(\mathbf{t}_a \times \mathbf{n_r})]_d \phi_k, \varphi_j \rangle_{\partial T}$ and the $\mathbb{L}^t$ submatrices by $\mathbb{L}_{ad,kj}^t = \langle \widehat{t}_{ad}\phi_k, \varphi_j \rangle_{\partial T}$. Finally, the subblocks of $\mathbb{M}^t$ are computed as $\mathbb{M}_{ab,k\ell}^t = \langle \nu_{ab}\phi_k, \phi_\ell \rangle_{\partial T}$, with $\nu_{ab} = \mathbf{t}_a^* \boldsymbol{G}\mathbf{t}_b$.

The linear form can be defined as $\mathbf{F}^t = [\mathbf{F}_1^t; \mathbf{F}_2^t]$, where the components of the vector are given by

$$
\mathbf{F}_{a,\ell}^t = \langle (\mathbf{t}_a \times \mathbf{n_r})^* \boldsymbol{G}\mathbf{e}_\partial, \phi_\ell \rangle_{\partial T \cap \partial \Omega_{\mathbf{r},E}} - \langle \widehat{\mathbf{t}}_a^* \mathbf{v}_\partial, \phi_\ell \rangle_{\partial T \cap \partial \Omega_{\mathbf{r},V}} - \langle \mathbf{t}_b^* \mathbf{f}_0, \phi_\ell \rangle_{\partial T \cap \partial \Omega_{\mathbf{r},\mathrm{rad}}}, \qquad \text{(A.7)}
$$

for $a = 1, 2$.

Table A.4
Non-affine functions for EIM.

Bilinear/Linear forms	Spatial points	Non-affine functions	# Non-affine terms
$\mathcal{B}$	Volume	$\boldsymbol{G}_{cd}$, $1 \le c \le d \le 3$	$\sum_{i=1}^{6} Q_i^{\mathcal{B}}$
$\mathcal{C}$	Face	γ_{cb}, $1 \le c \le 3$, $1 \le b \le 2$	$\sum_{i=1}^{6} Q_i^{\mathcal{C}}$
$\mathcal{E}$	Face	ϵ_{cb}, $1 \le c \le 3$, $1 \le b \le 2$	$\sum_{i=1}^{6} Q_i^{\mathcal{E}}$
$\mathcal{K}$	Face	κ_{cd}, $1 \le c, d \le 3$	$\sum_{i=1}^{9} Q_i^{\mathcal{K}}$
$\mathcal{D}$	Face	δ_{cd}, $1 \le c, d \le 3$	$\sum_{i=1}^{9} Q_i^{\mathcal{D}}$
$\mathcal{R}$	Face	$[\boldsymbol{G}(\mathbf{t}_a \times \mathbf{n_r})]_d$, $1 \le a \le 2$, $1 \le d \le 3$,	$\sum_{i=1}^{6} Q_i^{\mathcal{R}}$
$\mathcal{L}$	Face	$[\boldsymbol{G}\mathbf{t}_a]_d$, $1 \le a \le 2$, $1 \le d \le 3$	$\sum_{i=1}^{6} Q_i^{\mathcal{L}}$
$\mathcal{M}$	Face	$\mathbf{t}_a^* \boldsymbol{G}\mathbf{t}_b$, $1 \le a, b \le 2$	$\sum_{i=1}^{3} Q_i^{M}$
$\mathcal{F}_E$	E-tangent boundary	$(\mathbf{t}_a \times \mathbf{n_r})^* \boldsymbol{G}\mathbf{e}_\partial$, $1 \le a \le 2$	$\sum_{i=1}^{2} Q_i^{\mathcal{F}_E}$
$\mathcal{F}_V$	V-tangent boundary	$\mathbf{t}_a^* \boldsymbol{G}\mathbf{v}_\partial$, $1 \le a \le 2$	$\sum_{i=1}^{2} Q_i^{\mathcal{F}_V}$

Empirical interpolation

In order to apply the EIM to the elemental matrices in (A.1), we first need to identify the non-affine parametrized functions to be interpolated, as well as the discrete set of spatial points where the interpolants will be evaluated. In the finite element context, the natural choice is the Gaussian quadrature points $\mathbf{x}_\xi$ in the discretization, needed to compute the elemental inner products for volumes and faces. In addition, the non-affine functions can be readily identified from (A.2)–(A.7), and involve the components of the deformation tensor $\boldsymbol{G}$ for the volume bilinear forms and its effect on the normal and the tangent vectors for the face bilinear forms. Details of the non-affine functions are collected in Table A.4.

For every non-affine function $f(\mathbf{x}, \theta)$, the EIM also returns the Q spatial indices, denoted as $\mathcal{I}$, where the parameter-independent orthogonal basis $\boldsymbol{\Phi}$ needs to be evaluated in order to retrieve the coefficients $\boldsymbol{\alpha}(\theta)$ for a new θ, achieved through solving the $Q \times Q$ linear system

$$f(\mathbf{x}_\mathcal{I}, \theta) = \boldsymbol{\Phi}(\mathbf{x}_\mathcal{I})\boldsymbol{\alpha}(\theta).$$

Note that each non-affine function that is interpolated may require a different amount Q of orthogonal basis functions to meet the EIM accuracy requirements.

Once the non-affine functions are interpolated, we can approximate both $\mathcal{A}_q^\dagger$ and $\mathcal{F}_q^\dagger$ through affine expansions, namely

$$
\begin{aligned}
\mathcal{A}^\dagger(\mathbf{e}^\dagger, \boldsymbol{\xi}^\dagger; (\omega, \varepsilon, \theta)) \approx &\, \mathcal{A}(\mathbf{v}, \boldsymbol{\kappa}) - \sum_{i=1}^{6} \sum_{q=1}^{Q_i^{\mathcal{B}}} \alpha_{qi}^{\mathcal{B}}(\theta) \left[\mathcal{B}_{qi}(\mathbf{e}, \boldsymbol{\kappa}) + \mathcal{B}_{qi}(\mathbf{v}, \boldsymbol{\xi}) \right] \\
&- \omega^2 \mathcal{A}_\varepsilon(\mathbf{e}, \boldsymbol{\xi}) - \sum_{i=1}^{6} \sum_{q=1}^{Q_i^{\mathcal{C}}} \alpha_{qi}^{\mathcal{C}}(\theta) \mathcal{C}_{qi}(\widehat{\mathbf{e}}, \boldsymbol{\kappa}) \\
&+ \sum_{i=1}^{9} \sum_{q=1}^{Q_i^{\mathcal{K}}} \alpha_{qi}^{\mathcal{K}}(\theta) \mathcal{K}_{qi}(\mathbf{v}, \boldsymbol{\xi}) + \sum_{i=1}^{9} \sum_{q=1}^{Q_i^{\mathcal{D}}} \alpha_{qi}^{\mathcal{D}}(\theta) \mathcal{D}_{qi}(\mathbf{e}, \boldsymbol{\xi}) \\
&- \sum_{i=1}^{6} \sum_{q=1}^{Q_i^{\mathcal{E}}} \alpha_{qi}^{\mathcal{E}}(\theta) \mathcal{E}_{qi}(\widehat{\mathbf{e}}, \boldsymbol{\xi}) - \sum_{i=1}^{6} \sum_{q=1}^{Q_i^{\mathcal{R}}} \alpha_{qi}^{\mathcal{R}}(\theta) \mathcal{R}_{qi}(\mathbf{v}, \boldsymbol{\mu}) \\
&- \sum_{i=1}^{6} \sum_{q=1}^{Q_i^{\mathcal{L}}} \alpha_{qi}^{\mathcal{L}}(\theta) \mathcal{L}_{qi}(\mathbf{e}, \boldsymbol{\mu}) + \sum_{i=1}^{3} \sum_{q=1}^{Q_i^{\mathcal{M}}} \alpha_{qi}^{\mathcal{M}}(\theta) \mathcal{M}_{qi}(\widehat{\mathbf{e}}, \boldsymbol{\mu}) \\
\mathcal{F}^\dagger(\boldsymbol{\xi}^\dagger; (\omega, \varepsilon, \theta)) \approx &\, \sum_{i=1}^{2} \sum_{q=1}^{Q_i^{\mathcal{F}_E}} \alpha_{qi}^{\mathcal{F}_E}(\theta) \mathcal{F}_{E,qi}(\boldsymbol{\mu}) + \sum_{i=1}^{2} \sum_{q=1}^{Q_i^{\mathcal{F}_V}} \alpha_{qi}^{\mathcal{F}_V}(\theta) \mathcal{F}_{V,qi}(\boldsymbol{\mu}) \\
&+ \mathcal{F}_{\text{rad}}(\boldsymbol{\mu}; (\omega, \varepsilon)),
\end{aligned}
\tag{A.8}
$$

where the parameter-independent forms $\mathcal{B}_{qi}, \mathcal{C}_{qi}, \mathcal{D}_{qi}, \mathcal{E}_{qi}, \mathcal{K}_{qi}, \mathcal{R}_{qi}, \mathcal{L}_{qi}, \mathcal{M}_{qi}, \mathcal{F}_{E,qi}, \mathcal{F}_{V,qi}$ are the orthogonal spatial functions $\boldsymbol{\Phi}$ returned by the EIM. The expressions (A.8) are the detailed version of the more compact form given in (12).

Computational strategy

The linearity and recovered affine parametric dependence of the problem allow for an efficient offline-online decomposition strategy. The offline stage – parameter independent, computationally intensive but performed only once – comprises: (1) the computation of snapshots; (2) the POD compression that produces the orthonormalized snapshots $\zeta_n^\dagger, 1 \leq n \leq N_{\max}$ associated with the HDG approximation space at the selected parameter values; (3) the application of EIM for each bilinear form in Table A.4; and (4) the formation and storage of several parameter-independent small matrices and vectors $\mathcal{B}_{qi}, \mathcal{C}_{qi}, \mathcal{D}_{qi}, \mathcal{E}_{qi}, \mathcal{K}_{qi}, \mathcal{R}_{qi}, \mathcal{L}_{qi}, \mathcal{M}_{qi}, \mathcal{F}_{E,qi}, \mathcal{F}_{V,qi}$.

The online stage evaluates, for a new tuple $(\omega, \varepsilon, \theta)$, the non-affine functions at the EIM indices $\mathfrak{I}$ with a cost $\mathcal{O}(\mathcal{Q}\mathcal{N}_\mathfrak{I}^\dagger)$ and the EIM coefficients $\mathbf{c}(\theta)$ in $\mathcal{O}(2\mathcal{Q}^{\cdot 2})$ operations, where $\mathcal{Q}$ and $\mathcal{Q}^{\cdot 2}$ are given by

$$\mathcal{Q} = \sum_{i=1}^{6} \left[Q_i^{\mathcal{B}} + Q_i^{\mathcal{C}} + Q_i^{\mathcal{R}} + Q_i^{\mathcal{L}} \right] + \sum_{i=1}^{9} \left[Q_i^{\mathcal{D}} + Q_i^{\mathcal{K}} \right] + \sum_{i=1}^{3} Q_i^{\mathcal{M}},$$

$$\mathcal{Q}^{\cdot 2} = \sum_{i=1}^{6} \left[\left(Q_i^{\mathcal{B}}\right)^2 + \left(Q_i^{\mathcal{C}}\right)^2 + \left(Q_i^{\mathcal{R}}\right)^2 + \left(Q_i^{\mathcal{L}}\right)^2 \right] + \sum_{i=1}^{9} \left[\left(Q_i^{\mathcal{D}}\right)^2 + \left(Q_i^{\mathcal{K}}\right)^2 \right] + \sum_{i=1}^{3} \left(Q_i^{\mathcal{M}}\right)^2.$$

Evaluating the EIM functions has $\mathcal{N}^\dagger$-dependence. Nevertheless, the dimension of the set of EIM indices $\mathcal{N}_\mathfrak{I}^\dagger$ where the functions are evaluated is usually much smaller than the discrete set of Gaussian quadrature points $\mathbf{x}_\xi$, thus $\mathcal{N}_\mathfrak{I}^\dagger \ll \mathcal{N}^\dagger$. Computing the trial coefficients $\{\lambda_n^\dagger(\omega, \varepsilon)\}_{n=1}^{N}$ involves assembling the linear system in (A.8) with complexity $\mathcal{O}(\mathcal{Q}N^2)$ and its solution with complexity $\mathcal{O}(N^3)$, hence independent of the dimension $\mathcal{N}^\dagger$ of the HDG approximation space.

Finally, the dependence on $\mathcal{N}^\dagger$ also appears when computing $\mathbf{v}_N, \mathbf{e}_N$, with a complexity of $\mathcal{O}(4\mathcal{N}^\dagger N)$. In the ROM community this obstacle is avoided by never evaluating the field variables; if the quantity of interest is *linear* in the field variables, the offline-online strategy enables dropping the $\mathcal{N}^\dagger$ dependence in the online stage, see [36,45]. Unfortunately, the quantities of interest in plasmonic simulations, such as the field enhancement or the optical intensity, are in general nonlinear in the field variables, thus the complexity for the online stage becomes $\mathcal{N}^\dagger$-dependent. Despite this shortcoming, if the quantities of interest involve localized integrals –*e.g.* optical intensity through a surface surrounding the scatterer, or the field enhancement within sub-wavelength volumes– we can compute $\mathbf{v}_N, \mathbf{e}_N$ only for the required discretization elements, thus greatly reducing the online cost. The approximate electromagnetic fields $\mathbf{V}_N, \mathbf{E}_N$ are obtained applying (9) to $\mathbf{v}_N, \mathbf{e}_N$ elementwise.

In summary, the implications of the above strategy are twofold: first, if N and $\mathcal{Q}$ are small and the quantity of interest is localized, we shall achieve very fast output evaluation, usually several orders of magnitude faster than the HDG output; second, we may choose the HDG approximation very conservatively – to effectively eliminate the error between the exact output and HDG output – by only slightly affecting the online (marginal) cost.

Appendix B. Deformation mapping for coaxial nanogap structure

In this appendix, we introduce the deformation mapping that is employed for the periodic annular structure. For simplicity we present the mapping for only one ring of radius R and gap width w, although the extension to multiple rings is straightforward. We therefore have two parameters $\theta = (\theta_1, \theta_2)$ that specify the modified radius $R + \theta_1$ and gap width θ_2, see Fig. B.9. To derive the mapping, we resort to polar coordinates for the reference domain $(\rho_{\mathbf{r}}, \varphi_{\mathbf{r}})$ and the physical domain (ρ, φ). Furthermore, we have $\varphi_{\mathbf{r}} = \varphi$ since modifications occur only in the radial direction.

The physical coordinates relate to the reference coordinates as

$$x = \mathfrak{G}_1(\mathbf{x_r}) = \rho(\mathbf{x_r}) \cos\varphi(\mathbf{x_r}), \quad y = \mathfrak{G}_2(\mathbf{x_r}) = \rho(\mathbf{x_r}) \sin\varphi(\mathbf{x_r}), \quad z = \mathfrak{G}_3(\mathbf{x_r}) = z_{\mathbf{r}}.$$
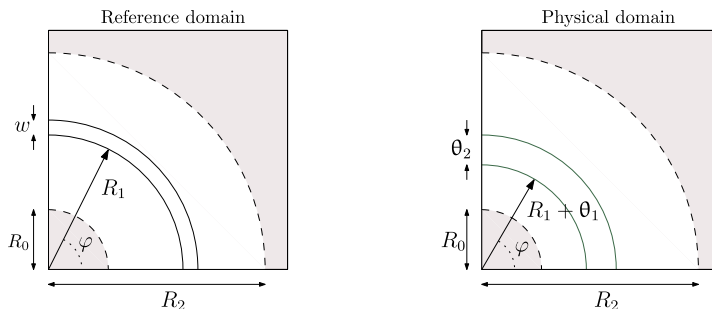

Fig. B.9. Reference and physical domain with relevant parameters.

In order to obtain $\rho(\mathbf{x_r})$ we will first compute $\rho(\rho_\mathbf{r}, z_\mathbf{r})$ and then apply the chain rule to compute the required derivatives. We propose to determine ρ using $\mathcal{C}^2$ cubic splines [12] in the radial direction, which enable us to parametrize the curved geometry exactly without introducing additional approximations. For both reference and physical domains in Fig. B.9, the gray zones correspond to regions that remain fixed, that is $\rho = \rho_\mathbf{r}$, thus avoiding the singularity at the origin.

We then use splines in the remaining regions that interpolate the knots

$$\{(\rho_{\mathbf{r},i}, \rho_i)\}_{i=1}^4 = \{(R_0, R_0),\ (R_1, R_1 + \theta_1),\ (R_1 + w, R_1 + \theta_1 + \theta_2),\ (R_2, R_2)\},$$

using cubic polynomials c_i for each pair $(\rho_{\mathbf{r},i-1}, \rho_{i-1})$ and $(\rho_{\mathbf{r},i}, \rho_i)$ such that $\rho = c_i(\rho_\mathbf{r})$, $i = 1, 2, 3$. To retrieve the coefficients of the cubics we impose continuity of their first and second derivatives (besides knot interpolation), rendering a spline that minimizes curvature. The extra condition to ensure a smooth blending with the gray zones is to prescribe unit slope at R_0, R_2, that is $c_1'(R_0) = c_3'(R_2) = 1$.

After solving a small 4x4 system and recovering the spline coefficients, the expression for the radial component can be compactly written as $\rho := \{c_i(\rho_{\mathbf{r},i})\}_{i=1}^4$. Nonetheless, this geometry modification is constant along the vertical direction, thus the mapping far from the scatterer (where radiation conditions are applied) is no longer the identity. One approach to circumvent this limitation is to use linear piecewise function $\mathcal{S}$ that equals one in the region of the film and gradually vanishes as we move away from it, and express the map $\rho(\rho_\mathbf{r}, z_\mathbf{r})$ as a convex combination of the spline and the identity, that is

$$\rho(\rho_\mathbf{r}, z_\mathbf{r}) := \mathcal{S}(z_\mathbf{r})\{c_i(\rho_{\mathbf{r},i})\}_{i=1}^4 + (1 - \mathcal{S}(z_\mathbf{r}))\rho_\mathbf{r}.$$

Finally, the required derivatives are computed invoking the chain rule, for instance

$$\frac{\partial x}{\partial x_\mathbf{r}} = \frac{\partial \rho}{\partial \rho_\mathbf{r}}\frac{\partial \rho_\mathbf{r}}{\partial x_\mathbf{r}}\cos\varphi + \rho\frac{\partial\cos\varphi}{\partial x_\mathbf{r}} = \left[\mathcal{S}(z_\mathbf{r})\{c_i'(\rho_{\mathbf{r},i})\}_{i=1}^4 + 1 - \mathcal{S}(z_\mathbf{r})\right]\frac{x_\mathbf{r}}{\rho_\mathbf{r}}\cos\varphi + \rho\sin\varphi\frac{y_r}{\rho_\mathbf{r}}.$$

It is then immediate to compute the Jacobian $\mathcal{G}_{ij} = \partial_{\mathbf{x}_{\mathbf{r},j}}\mathfrak{G}_i(\mathbf{x_r})$, its determinant g and the required symmetric tensor $\boldsymbol{G} = g^{-1}\mathcal{G}^T\mathcal{G}$.

References

[1] M.N. Albunni, V. Rischmuller, T. Fritzsche, B. Lohmann, Model-order reduction of moving nonlinear electromagnetic devices, IEEE Trans. Magn. 44 (7) (2008) 1822–1829.

[2] M. Barrault, Y. Maday, N.C. Nguyen, A.T. Patera, An 'empirical interpolation' method: application to efficient reduced-basis discretization of partial differential equations, C. R. Math. 339 (9) (2004) 667–672.

[3] J.-P. Berenger, A perfectly matched layer for the absorption of electromagnetic waves, J. Comput. Phys. 114 (2) (1994) 185–200.

[4] G. Berkooz, P. Holmes, J.L. Lumley, The proper orthogonal decomposition in the analysis of turbulent flows, Annu. Rev. Fluid Mech. 25 (1) (1993) 539–575.

[5] M.L. Brongersma, Plasmonic photodetectors, photovoltaics, and hot-electron devices, Proc. IEEE 104 (12) (2016) 2349–2361.

[6] S. Burgard, O. Farle, R. Dyczij-Edlinger, A novel parametric model order reduction approach with applications to geometrically parameterized microwave devices, COMPEL: Int. J. Comput. Math. Electr. Electron. Eng. 32 (5) (2013) 1525–1538.

[7] S. Burgard, A. Sommer, O. Farle, R. Dyczij-Edlinger, Reduced-order models of finite-element systems featuring shape and material parameters, Electromagnetics 34 (3–4) (2014) 143–160.

[8] S. Chaturantabut, D.C. Sorensen, Nonlinear model reduction via discrete empirical interpolation, SIAM J. Sci. Comput. 32 (5) (2010) 2737–2764.

[9] Y. Chen, J.S. Hesthaven, Y. Maday, A seamless reduced basis element method for 2d Maxwell's problem: an introduction, in: Spectral and High Order Methods for Partial Differential Equations, Springer, 2011, pp. 141–152.

[10] Y. Chen, J.S. Hesthaven, Y. Maday, J. Rodríguez, Certified reduced basis methods and output bounds for the harmonic Maxwell's equations, SIAM J. Sci. Comput. 32 (2) (2010) 970–996.

[11] B. Cockburn, B. Dong, J. Guzmán, A superconvergent LDG-hybridizable Galerkin method for second-order elliptic problems, Math. Comput. 77 (264) (2008) 1887–1916.

[12] C. De Boor, C. De Boor, E.-U. Mathématicien, C. De Boor, C. De Boor, A Practical Guide to Splines, vol. 27, Springer-Verlag, New York, 1978.

[13] V. De La Rubia, U. Razafison, Y. Maday, Reliable fast frequency sweep for microwave devices via the reduced-basis method, IEEE Trans. Microw. Theory Tech. 57 (12) (2009) 2923–2937.

[14] P. Drude, Zur elektronentheorie der metalle, Ann. Phys. 306 (3) (1900) 566–613.

[15] R. Dyczij-Edlinger, O. Farle, Finite element analysis of linear boundary value problems with geometrical parameters, COMPEL: Int. J. Comput. Math. Electr. Electron. Eng. 28 (4) (2009) 779–794.

[16] M. Fares, J.S. Hesthaven, Y. Maday, B. Stamm, The reduced basis method for the electric field integral equation, J. Comput. Phys. 230 (14) (2011) 5532–5555.

[17] J. Fink, W. Rheinboldt, On the error behavior of the reduced basis technique for nonlinear finite element approximations, ZAMM Z. Angew. Math. Mech. (J. Appl. Math. Mech.) 63 (1) (1983) 21–28.

[18] M. Ganesh, J.S. Hesthaven, B. Stamm, A reduced basis method for electromagnetic scattering by multiple particles in three dimensions, J. Comput. Phys. 231 (23) (2012) 7756–7779.

[19] M.A. Grepl, Y. Maday, N.C. Nguyen, A.T. Patera, Efficient reduced-basis treatment of nonaffine and nonlinear partial differential equations, ESAIM: Math. Model. Numer. Anal. 41 (03) (2007) 575–605.

[20] D. Grischkowsky, S. Keiding, M. Van Exter, C. Fattinger, Far-infrared time-domain spectroscopy with terahertz beams of dielectrics and semiconductors, JOSA B 7 (10) (1990) 2006–2015.

[21] M. Groner, J. Elam, F. Fabreguette, S.M. George, Electrical characterization of thin Al_2O_3 films grown by atomic layer deposition on silicon and various metal substrates, Thin Solid Films 413 (1) (2002) 186–197.

[22] M. Hammerschmidt, C. Barth, J. Pomplun, S. Burger, C. Becker, F. Schmidt, Reconstruction of photonic crystal geometries using a reduced basis method for nonlinear outputs, in: Photonic and Phononic Properties of Engineered Nanostructures VI, vol. 9756, International Society for Optics and Photonics, 2016, p. 97561R.

[23] M.W. Hess, P. Benner, Fast evaluation of time-harmonic Maxwell's equations using the reduced basis method, IEEE Trans. Microw. Theory Tech. 61 (6) (2013) 2265–2274.

[24] J.S. Hesthaven, G. Rozza, B. Stamm, et al., Certified Reduced Basis Methods for Parametrized Partial Differential Equations, Springer, 2016.

[25] J.S. Hesthaven, B. Stamm, S. Zhang, Certified reduced basis method for the electric field integral equation, SIAM J. Sci. Comput. 34 (3) (2012) A1777–A1799.

[26] A. Huerta, A. Angeloski, X. Roca, J. Peraire, Efficiency of high-order elements for continuous and discontinuous Galerkin methods, Int. J. Numer. Methods Eng. 96 (9) (2013) 529–560.

[27] G. Kerschen, J.-c. Golinval, A.F. Vakakis, L.A. Bergman, The method of proper orthogonal decomposition for dynamical characterization and order reduction of mechanical systems: an overview, Nonlinear Dyn. 41 (1–3) (2005) 147–169.

[28] K. Kirchner, K. Urban, O. Zeeb, Maxwell's equations for conductors with impedance boundary conditions: discontinuous Galerkin and reduced basis methods, ESAIM: Math. Model. Numer. Anal. 50 (6) (2016) 1763–1787.

[29] Y. Maday, A.T. Patera, G. Turinici, Global a priori convergence theory for reduced-basis approximations of single-parameter symmetric coercive elliptic partial differential equations, C. R. Math. 335 (3) (2002) 289–294.

[30] S.A. Maier, Plasmonics: Fundamentals and Applications, Springer Science & Business Media, 2007.

[31] S.A. Maier, H.A. Atwater, Plasmonics: localization and guiding of electromagnetic energy in metal/dielectric structures, J. Appl. Phys. 98 (1) (2005) 10.

[32] G. Mur, Absorbing boundary conditions for the finite-difference approximation of the time-domain electromagnetic-field equations, IEEE Trans. Electromagn. Compat. 4 (1981) 377–382.

[33] N.C. Nguyen, J. Peraire, B. Cockburn, An implicit high-order hybridizable discontinuous Galerkin method for linear convection–diffusion equations, J. Comput. Phys. 228 (9) (2009) 3232–3254.

[34] N.C. Nguyen, J. Peraire, B. Cockburn, High-order implicit hybridizable discontinuous Galerkin methods for acoustics and elastodynamics, J. Comput. Phys. 230 (10) (2011) 3695–3718.

[35] N.C. Nguyen, J. Peraire, B. Cockburn, Hybridizable discontinuous Galerkin methods for the time-harmonic Maxwell's equations, J. Comput. Phys. 230 (19) (2011) 7151–7175.

[36] N.C. Nguyen, K. Veroy, A.T. Patera, Certified real-time solution of parametrized partial differential equations, in: Handbook of Materials Modeling, Springer, 2005, pp. 1529–1564.

[37] A.K. Noor, J.M. Peters, Reduced basis technique for nonlinear analysis of structures, AIAA J. 18 (4) (1980) 455–462.

[38] L. Novotny, N. Van Hulst, Antennas for light, Nat. Photonics 5 (2) (2011) 83–90.

[39] M.A. Ordal, R.J. Bell, R.W. Alexander, L.L. Long, M.R. Querry, Optical properties of fourteen metals in the infrared and far infrared: Al, Co, Cu, Au, Fe, Pb, Mo, Ni, Pd, Pt, Ag, Ti, V, and W, Appl. Opt. 24 (24) (1985) 4493–4499.

[40] E. Ozbay, Plasmonics: merging photonics and electronics at nanoscale dimensions, Science 311 (5758) (2006) 189–193.

[41] E.D. Palik, Handbook of Optical Constants of Solids, vol. 3, Academic Press, 1998.

[42] H.-R. Park, X. Chen, N.C. Nguyen, J. Peraire, S.-H. Oh, Nanogap-enhanced terahertz sensing of 1 nm thick ($\lambda/10^6$) dielectric films, ACS Photonics 2 (3) (2015) 417–424.

[43] P.-O. Persson, J. Bonet, J. Peraire, Discontinuous Galerkin solution of the Navier–Stokes equations on deformable domains, Comput. Methods Appl. Mech. Eng. 198 (17) (2009) 1585–1595.

[44] J. Pomplun, F. Schmidt, Reduced basis method for electromagnetic field computations, in: Scientific Computing in Electrical Engineering SCEE 2008, Springer, 2010, pp. 85–92.

[45] C. Prud'homme, D.V. Rovas, K. Veroy, L. Machiels, Y. Maday, A.T. Patera, G. Turinici, Reliable real-time solution of parametrized partial differential equations: reduced-basis output bound methods, J. Fluids Eng. 124 (1) (2002) 70–80.

[46] A. Quarteroni, A. Manzoni, F. Negri, Reduced Basis Methods for Partial Differential Equations: An Introduction, vol. 92, Springer, 2015.

[47] A. Quarteroni, G. Rozza, A. Manzoni, Certified reduced basis approximation for parametrized partial differential equations and applications, J. Math. Ind. 1 (1) (2011) 3.

[48] H. Raether, Excitation of Plasmons and Interband Transitions by Electrons, Springer, 1980.

[49] G. Rozza, D.B.P. Huynh, A.T. Patera, Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations, Arch. Comput. Methods Eng. 15 (3) (2007) 1.

[50] J. Saà-Seoane, N.C. Nguyen, H. Men, R. Freund, J. Peraire, Binary optimization techniques for linear PDE-governed material design, Appl. Phys. A 109 (4) (2012) 1023–1030.

[51] L. Sirovich, Turbulence and the dynamics of coherent structures. I—Coherent structures. II—Symmetries and transformations. III—Dynamics and scaling, Q. Appl. Math. 45 (1987) 561–571.

[52] I.M. Sobol, On the distribution of points in a cube and the approximate evaluation of integrals, Ž. Vyčisl. Mat. Mat. Fiz. 7 (4) (1967) 784–802.

[53] A. Sommerfeld, Partial Differential Equations in Physics, vol. 1, Academic Press, 1949.

[54] B. Špačková, P. Wrobel, M. Bocková, J. Homola, Optical biosensors based on plasmonic nanostructures: a review, Proc. IEEE 104 (12) (2016) 2380–2408.

[55] F. Vidal-Codina, N.-C. Nguyen, S.-H. Oh, J. Peraire, A hybridizable discontinuous Galerkin method for computing nonlocal electromagnetic effects in three-dimensional metallic nanostructures, J. Comput. Phys. 355 (2018) 548–565.

[56] S. Volkwein, Model reduction using proper orthogonal decomposition, Lecture Notes, Institute of Mathematics and Scientific Computing, University of Graz, see http://www.uni-graz.at/imawww/volkwein/POD.pdf, 2011.

[57] M.W. Hess, P. Benner, A reduced basis method for microwave semiconductor devices with geometric variations, COMPEL: Int. J. Comput. Math. Electr. Electron. Eng. 33 (4) (2014) 1071–1081.

[58] K. Willcox, J. Peraire, J. White, An Arnoldi approach for generation of reduced-order models for turbomachinery, Comput. Fluids 31 (3) (2002) 369–389.

[59] D. Yoo, N.C. Nguyen, L. Martin-Moreno, D.A. Mohr, S. Carretero-Palacios, J. Shaver, J. Peraire, T.W. Ebbesen, S.-H. Oh, High-throughput fabrication of resonant metamaterials with ultrasmall coaxial apertures via atomic layer lithography, Nano Lett. 16 (3) (2016) 2040–2046.