Simulation methods for plasmonic structures

by

Ferran Vidal-Codina

Llicenciatura de Matemàtiques, UPC BarcelonaTech (2010)
Enginyeria de Camins Canals i Ports, UPC BarcelonaTech (2011)
S.M., Computation for Design and Optimization, MIT (2013)

Submitted to the Department of Aeronautics and Astronautics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Computational Science and Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2017

Author .....
Department of Aeronautics and Astronautics
May 25, 2017

Certified by .....
Thesis Supervisor: Jaume Peraire
H. N. Slater Professor of Aeronautics and Astronautics

Certified by .....
Thesis Supervisor: Ngoc-Cuong Nguyen
Principal Research Scientist

Certified by .....
Thesis Committee Member: Youssef M. Marzouk
Associate Professor of Aeronautics and Astronautics

Certified by .....
Thesis Committee Member: Anthony T. Patera
Ford Professor of Mechanical Engineering

Accepted by .....
Nicolas G. Hadjiconstantinou
Professor of Mechanical Engineering
Co-Director, Computational Science and Engineering

Accepted by .....
Youssef M. Marzouk
Associate Professor of Aeronautics and Astronautics
Chair, Graduate Program Committee

Simulation methods for plasmonic structures

by

Ferran Vidal-Codina

Submitted to the Department of Aeronautics and Astronautics
on May 25, 2017, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Computational Science and Engineering

Abstract

In the recent years there has been a growing interest in studying electromagnetic wave propagation at the nanoscale. The interaction of light with metallic nanostructures produces a collective excitation of conduction electrons at the metal surface, also known as surface plasmons. These plasmonic resonances enable an unprecedented control of light by confining the electromagnetic field to regions well beyond the diffraction limit, thereby leading to near-field enhancements of the incident wave of several orders of magnitude. These remarkable properties have motivated the application of plasmonic devices in sensing, nano-resolution imaging, energy harvesting, nanoscale electronics and cancer treatment.

Despite state-of-the-art nanofabrication techniques are used to realize plasmonic devices, their performance is severely impacted by fabrication uncertainties arising from extreme manufacturing constraints. Mathematical modeling and numerical simulation are therefore essential to accurately predict the response of the physical system, and must be incorporated in the design process. Nonetheless, plasmonic simulations present notable challenges. From the physical perspective, the realistic behavior of conduction electrons in metallic nanostructures is not captured by Maxwell's equations, thus requiring additional modeling. From the simulation perspective, the disparity in length scales stemming from the extreme field localization exceeds the capabilities of most numerical simulation schemes. In addition, relevant data such as optical constants or geometry specifications are typically subject to measurement and manufacturing errors, hence simulations need to accommodate uncertainty in the data.

In this thesis we present a collection of numerical methods to efficiently simulate electromagnetic wave propagation through metallic nanostructures. Firstly, we develop the hybridizable discontinuous Galerkin (HDG) method for Maxwell's equations augmented with the hydrodynamic model for metals, which accounts for the nonlocal interactions between electrons that become predominant at nanometric regimes. Secondly, we develop a reduced order modeling (ROM) framework for Maxwell's equations with the HDG method, enabling the incorporation of material and geometric uncertainties in the simulations. The result is a family of surrogate models that produces accurate yet inexpensive simulations of plasmonic devices. Finally, we apply these approaches to the study of periodic annular nanogaps, and present parametric analyses, verification with experimental data and design of novel structures.

Thesis Supervisor: Jaume Peraire
Title: H. N. Slater Professor of Aeronautics and Astronautics

Thesis Supervisor: Ngoc-Cuong Nguyen
Title: Principal Research Scientist

Thesis Committee Member: Youssef M. Marzouk
Title: Associate Professor of Aeronautics and Astronautics

Thesis Committee Member: Anthony T. Patera
Title: Ford Professor of Mechanical Engineering

Acknowledgments

First and foremost I want to express my gratitude to Prof. Jaume Peraire, my advisor during the PhD. I feel indebted to him for giving me the extraordinary opportunity to come to MIT. His profound knowledge and intuition have been very valuable whenever I could not find answers to the problems. The guidance, support, mentorship and advice he provided me has definitely made this work possible. In addition, having an advisor with whom you can regularly discuss football, politics and even play squash games has indeed made this experience as enriching as it can get. It has truly been a pleasure working with him. I am also very grateful to Dr. Cuong Nguyen, who has also supervised my work and has helped me through his vast technical knowledge. I have greatly enjoyed working with such a talented researcher. I would also like to thank the other members of my thesis committee, Prof. Youssef Marzouk and Prof. Anthony Patera, for contributing to this work with valuable suggestions, ideas and discussions, as well as providing useful feedback in every step of the way. I would also like to highlight the assistance of Prof. Sang-Hyun Oh for providing relevant applications and research directions, as well as his continued interest in my work. I would also like to mention the effort of Prof. Nicholas Fang by greatly improving the quality of the thesis. Finally, I am proud to have had the opportunity to meet and work with Dr. Xevi Roca. His enormous and continuous help, assistance, expertise and friendship have been invaluable to me and to this work. I would also like to recognize Prof. Antonio Huerta and Prof. Rosa Maria Estela, who mentored me through college, and without whom I would have never taken this step forward in my life.

I want to acknowledge 'Obra Social La Caixa' for giving me the once in a lifetime opportunity to go to gradschool in the United States through one of their generous fellowships, as well as AFOSR Grant No. FA9550-11-1-0141, FA9550-12-0357 and the Singapore MIT-Alliance for funding this research.

I feel extremely thankful for having shared these six years (or some part thereof) with wonderful people, who have definitely made the day-to-day much more enjoyable. My acknowledgements go to my fellow PhD students and friends Joel, Xevi, David, Carmen, Hemant and Pablo for being there in all the great moments and lifting me up during difficult times. This journey would not have been possible without you, and these last two years I have

greatly missed not having you around as often as I would have liked. I also thank my fellow ACDLers Patrick, Phil, Antoni, Remi, Alessio, Andy, Eric, Victor, Hugh and Max for their support and friendship, as well as Jean, Joyce, Anthony, Robin, Beth and Kate for helping me deal with bureaucracy always with a smile on their faces.

Outside MIT, I have also had the immense fortune of meeting multiple friends throughout my years, with whom I have shared plenty of experiences that I will never forget. Among them, I would like to recognize Paula, Enrique, Toni, Àlex, Maite, Noel, Carlos, Íñigo, Adrià, Helena, Juanito, Quique, Hasier, Ivan, together with all the people within the Spain@MIT football team for all the successes we achieved together. In addition, I am extremely grateful to Pipo, Marc, Margot, Judit, Diana, Inés, Ferran and Davids for turning the toughest year into an unbelievable one. Finally, I want to thank Mariona for partly sharing this exciting journey, and for making me the better person that I am today.

Last but definitely not least, I have to thank the many friends back in Catalunya that have always made me feel as if I had not left at all. My deepest gratitude goes to Pol, Marc, Oleguer, Serra, Rafi, Darnés, Gil, Faja, Sala, Dasz, Sito, Eloi and Cano, along with Pau, Marta, Dr. Ràpid, Guille, Juli, Sergi, Gali and Carla. I would like to specially acknowledge Membri and Valley for their advice and friendship through difficult times. It will certainly never be forgotten.

My family has been of great importance since the beginning, and I could not have possibly made it that far without their constant support and understanding. Knowing that Oriol, Marta, Joan, Nil, Raül, my uncles and grandparents always had my back made overcoming any obstacle an easy task. To conclude, I would like to specially dedicate this work to my mum Maria Carme, my dad Carles and my brother Adrià, for being the best family one can dream of, and for teaching me the most useful and humane lessons I have learned in life. Ben amunt per recordar d'on venim, ben amunt per recordar qui som i ben amunt per recordar qui volem ser.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The field of plasmonics [153, 155, 196, 234] studies the collective excitation of conduction-band electrons that occurs due to the interaction of light with metallic nanostructures. These resonances can be classified as localized surface plasmons or surface plasmon polaritons. Localized surface plasmons are non-propagating excitations that appear naturally in the surface of nanoparticles when illuminated by an electromagnetic (EM) wave. Conversely, surface plasmon polaritons are dispersive electromagnetic waves that propagate at the interface between a metal and a dielectric, excited by the coupling of the incident EM wave with the metal's electron plasma. Plasmon resonances are able to squeeze light in volumes several orders of magnitude smaller than the wavelength of light, leading to large near-field enhancements of the incident EM wave. The excitation of surface plasmons is magnified near the corners or singularities of metallic nanoparticles, or within gaps formed by metallic structures at the nanoscale. Moreover, the extreme confinement and enhancement properties provide unparalleled means for manipulation of light and its interaction with metals, well beyond the diffraction limit. As a result, plasmonics has motivated applications in sensing [7, 27, 33, 106, 124, 133, 141, 142, 243], energy harvesting [12, 28, 34, 76, 77], near-field scanning microscopy [80,99,107,125,126,186,267,274], plasmonic waveguiding, amplification and lasing [18,26,154,195,196,202,238,272] , optical data storage [35,156,187,245,280] and cancer theranostics [58, 62, 66, 104–106].

In order to illustrate the trapping of light in deep-subwavelength apertures, we consider the interaction of an electromagnetic wave with two infinitely long gold nanowires of radius 25

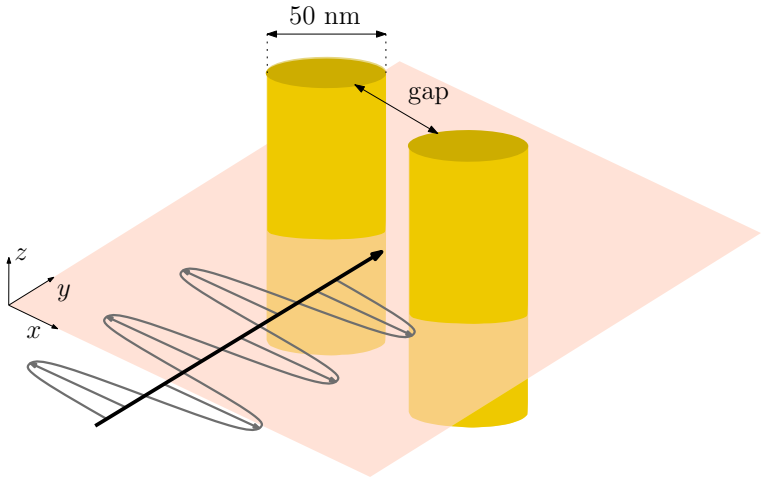


Figure 1-1: Schematic diagram of nanowire dimer illuminated with electromagnetic wave.

nm in free space, separated by a small gap, see Fig. 1-1. This plasmonic device is commonly known as plasmonic dimer, and it has been shown to support strong field enhancements in the gap region [96,185] for nanometric and subnanometric separation distances. The dimer is illuminated with an x-polarized plane wave that propagates in the y direction, exciting a localized surface plasmon as the gap decreases. The photon energy of the incoming wave is fixed at 3 eV, which corresponds to a wavelength of 413 nm, that is 16 times greater than the radius of the wires. To demonstrate the confinement properties of the dimer, the absolute value of the electric field's x-component is shown in Fig. 1-2 for several values of the gap-radius ratio. Initially, for a separation comparable to the radius, the enhancement is minimal and not even concentrated in the gap region, since the interaction between the nanowires is poor. Nonetheless, as the gap shrinks the growing collective oscillation of electrons at the metal surface is able to squeeze the incident light in the volume between the wires. The enhancement caused by the plasmon resonance is greater for smaller separations, reaching almost three orders of magnitude for a 5 Å gap – almost 1000 times smaller than the wavelength.

The plasmonic devices of interest in this dissertation are periodic annular gap structures, which have been shown to produce extraordinary optical transmission and enormous field enhancements [14, 63, 71, 114, 205, 218]. These structures consist of periodic arrays of sub-wavelength annular apertures of a dielectric material patterned in a metallic film, and unlike arrays of circular and rectangular apertures they sustain plasmon resonances for a broad range of frequency regimes. That is, for a fixed aperture size one can adjust the ring di-

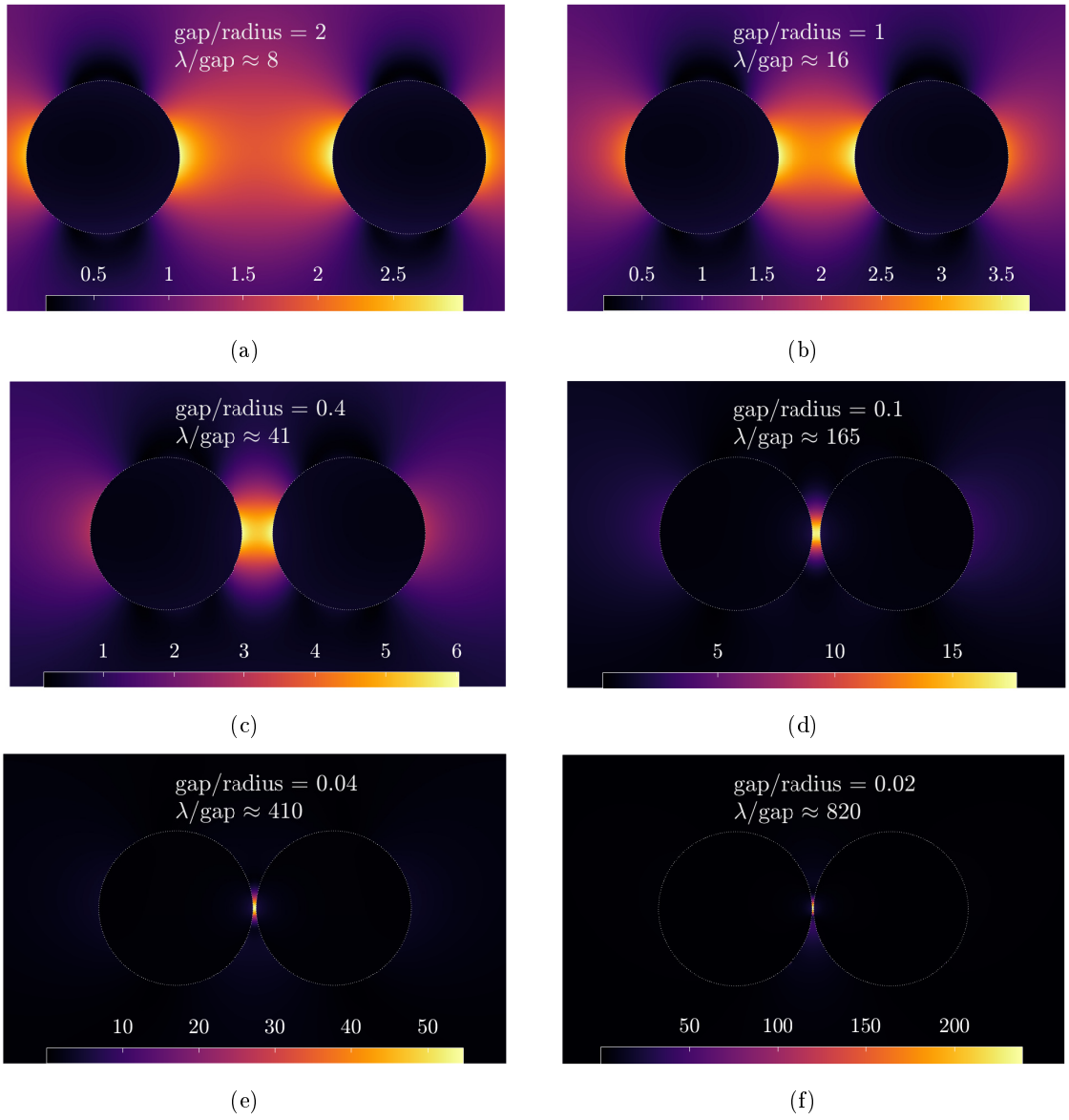


Figure 1-2: x-component of absolute value of electric field for 25 nm radius gold dimer (highlighted with white dots) at several gap separations, for constant incident photon energy of 3 eV.

ameter and the array periodicity to generate resonances for the visible, the mid infrared (MIR) and the far infrared (FIR) regime. Due to this extended frequency range of operations, annular gap structures have numerous applications such as sensing in low THz frequencies [188, 189, 198], optical trapping [12, 226, 252] or nonlinear optics [9].

Plasmonic phenomena are governed by the propagation of electromagnetic waves, described by Maxwell's equations. These waves propagate through dielectric as well as metallic media, hence several models have been proposed to characterize the behavior of metallic nanostruc-

tures. The research presented in this thesis is devoted to high-fidelity simulation techniques for 2d and 3d plasmonic structures, with a special focus on annular gap structures.

1.1 Challenges and scope

The ability to accurately model and simulate electromagnetic wave propagation problems for plasmonic applications requires capabilities that challenge traditional modeling and simulation techniques. Firstly, the interaction of long-wavelength light (μm and mm) with nanometric cavities involves an enormous mismatch in relevant length scales, see for instance Figs. 1-2e and 1-2f. This circumstance can quickly overcome the capabilities of conventional finite difference and finite volumes schemes, thus it is crucial to develop advanced simulation methods that are capable of resolving multiple length scales.

Plasmonic devices often present curved geometries. To compute high-fidelity results an accurate representation of the geometry is paramount. Additionally, plasmonic phenomena are characterized by the extreme confinement and tight localization of fields in nanometer-wide apertures. Consequently, the discretizations required to attain faithful simulations need to be adaptive (to focalize the degrees of freedom in the regions of interest) and anisotropic (to properly capture boundary-layer type structures that appear at the interface of metallic nanostructures.)

Maxwell's equations are non-coercive, hence leading to indefinite linear systems of equations after an appropriate discretization of the computational domain and the equations. This circumstance precludes the use of iterative methods to compute the high-fidelity solutions, and the development of preconditioners for Maxwell's equations is still an active area of research [6, 85, 121, 253]. Also, the use of iterative methods for metal-dielectric systems at the nanoscale is an unexplored problem. Direct solutions to the linear problem are computationally demanding, specially for 3d structures. Moreover, plasmonic devices exhibit an enormous contrast of optical constants between metals and dielectrics. Hence, to tackle 3d plasmonic simulations we not only need significant computational power and storage, but also numerical schemes that can address the aforementioned issues.

From a manufacturing perspective the fabrication of nanometer-wide gap structures is difficult, and small geometric deviations from the nominal design may produce entirely different

responses of the plasmonic device. Furthermore, the material constants that characterize the optical response of metals and dielectrics are determined experimentally, thus subject to measurement errors. If the ultimate goal is to endow the numerical simulation methods with a predictive power, it is of foremost importance to incorporate fabrication tolerances as well as measurement uncertainties in the simulation and design process.

In terms of modeling, the most common technique to simulate plasmonic devices is to solve Maxwell's equations in both the metal and the dielectric, and use a permittivity in the metal given by Drude's model. This approach is known as local response approximation (LRA). This model, albeit simple and directly applicable to most numerical schemes, is limited, due to simplifications in the description of the electron motion. The limitations of LRA to properly model plasmonic resonances appear at the nanometer scale, where effects such as nonlocal interactions between electrons or quantum phenomena become non-negligible. To account for these effects, the mathematical model must be augmented with additional equations that capture the physics. In this thesis, we consider the hydrodynamic model (HM) for noble metals, which models the inter-electron coupling by including a hydrodynamic pressure term that is simultaenously solved with Maxwell's equations. The HM has been demonstrated to produce more physically meaningful results for plasmonic structures with noble metals below tenths of nanometers, compared to the LRA, even though it neglects quantum effects. In terms of numerical computation, the new field variables introduced by the additional equations increase the computational complexity of the simulations. Furthermore, the hydrodynamic model excites subnanometric phenomena omitted by LRA, thus we often require greater spatial resolution to attain accurate solutions.

1.2 Proposed methodology

The main contribution of this dissertation is a high-order numerical scheme, the hybridizable discontinuous Galerkin (HDG) method, to simulate the interaction of light with metallic nanostructures. The unique features of the HDG method make it particularly attractive for electromagnetic wave propagation problems: (1) it can be used on general unstructured meshes, thus allowing complex geometries and facilitating the use of adaptive discretizations; (2) it is high-order accurate, low dissipative and dispersive, thus particularly suited for

wave propagation problems; (3) the linear system that needs to be solved comprises only a reduced number of degrees of freedom, defined on the faces of the discretization cells; (4) the treatment of boundary conditions is naturally incorporated in the weak formulation by approximating the trace of the solution on the boundaries; (5) we need not consider special approximation spaces, since the discontinous Galerkin spaces include the commonly used curl-conforming subspaces that contain the exact solution; and (6) the discontinuous nature and stabilization techniques can easily accommodate material contrasts at the interfaces of several orders of magnitude.

The HDG method is developed for the LRA and for the HM for metals to account for nonlocal interactions between electrons. Moreover, since we solve Maxwell's equations in the dielectric and the Maxwell's equations augmented with hydrodynamic model in the metal, appropriate coupling conditions need to be devised.

In addition, we develop a reduced order model (ROM) framework to account for fabrication tolerances and uncertainty in the material properties within the LRA model. To that end, we develop a reference domain formulation for the time-harmonic Maxwell's equations that, along with empirical interpolation techniques, allows us to describe geometry deformations in terms of affine parametric expressions. Furthermore, an efficient implementation of ROM based on an offline-online computational strategy enables the computation of approximate solutions of large 3d problems in real-time. Hence, we believe the ROM techniques proposed in this thesis can be of tremendous interest for the design of novel plasmonic devices, due to the possibility of exploring multiple geometry configurations at a fraction of the cost of the full HDG simulation.

1.3 Literature review

In this section, we provide a literature review of the relevant approaches to simulate wave propagation phenomena, as well as the construction of reduced order models. The review of physical models, their limitations and applicability is deferred to Chapter 2.

1.3.1 Wave simulation techniques

The simulation of electromagnetic wave propagation phenomena has been a very active field of research in recent years, and multiple simulation approaches may be found in literature, both for the time and frequency domain formulation of Maxwell's equations. The two main families of methods for computational electromagnetics are integral equation models and differential equation models, which constitute two very distinct approaches to simulate wave propagation. For detailed reviews of the distinct methods for computational electromagnetics we refer the reader to [159, 223] and the references therein.

Integral equation models

The basic idea of integral equation models is the use of Green's functions, whereby we can solve for the volume integrals of the induced sources located on the boundary of the volume of interest, rather than solving for the entire unknown field in the volume. Hence, integral equation models enable the reduction of a 3d problem to a 2d problem over the boundary. In addition, Green's functions naturally account for unbounded domains, thus no artificial boundaries must be devised.

Unfortunately, integral equation models give rise to dense systems of equations, which albeit smaller –since they represent a 2d discretization– are computationally demanding to solve. Moreover, these methods are restricted for cases where Green's functions are available, thus limiting the applicability of integral equation models for heterogeneous materials or materials with space-varying properties. Finally, the treatment of inhomogeneities, anisotropy or nonlinearities is not straightforward. Therefore other approaches are preferred for complex systems and devices.

Among integral equation models the family of boundary element methods, also known as the method of moments in electromagnetics [82, 98, 209, 236], is the most extensively used.

Differential equation models

The family of differential equation models differs from integral equation models in that it pursues a discretization of Maxwell's equations, and the objective is to compute the value

of the unknown fields in the degrees of freedom on the discretized volume of interest.

The first and most widely used method for computational electromagnetics is the finite-difference time domain (FDTD) algorithm [132,246,247], which discretizes both space and time using Yee's scheme [276]. The main advantage of Yee's scheme is its simplicity and efficiency, due to the use of staggered Cartesian grids and second-order schemes for both space and time to simulate the evolution of Maxwell's equations in a fully explicit manner. The main limitation of FDTD is their extension to complex geometries with sharp or curved features, since Cartesian grids can only approximate these irregular boundaries in a staircased manner. If accurate simulations are sought, in order to avoid numerical diffraction a very high discretization resolution is needed, thus increasing the computational requirement. This is even more dramatic for plasmonic applications due to the extreme mismatch in length scales, limiting the applicability of FDTD. For photonic crystal simulations we should highlight MEEP [194], an open-source implementation of FDTD, which provides a robust, scalable and parallel framework to simulate the interaction of electromagnetic waves with photonic crystals, whereas Lumerical [2] is a software designed for the 2d and 3d FDTD simulation of surface plasmons.

Finite element (FE) methods [118, 160, 275] are amongst the most popular techniques for wave propagation problems, thanks to their ability to handle heterogeneous media and complex geometries with the use of unstructured grids. The class of face/edge elements introduced by Raviart-Thomas and Nédélec [23,167,168,210] have been extensively used to simulate electromagnetic wave propagation, and have been shown to avoid the problem of spurious modes [24] by explicitly enforcing the divergence condition. Edge elements can be further subdivided in low order elements, also known as Whitney elements [269], and high order and h/p elements [3,4,56,57,232,268]. High order edge elements are often desirable due to the reduced number of elements needed to capture the solutions compared to Whitney elements, specially for high frequencies, although the linear systems that stem from the discretization become denser with the approximation order of the elements. A commonly used implementation of edge elements for Maxwell's equations is the one provided by the RF Module of Comsol Multiphysics [1], which has been extended to include the hydrodynamic model [250]. Additionally, a frequency-domain implementation of the hydrodynamic model based on edge elements has been applied to the numerical simulation of 2d grooves and

nanowires [102].

An attractive alternative to edge elements is the class of discontinous Galerkin (DG) methods [13, 16, 53], since it leads to methods that are unstructured, locally conservative, high-order accurate, low dissipative and dispersive, optimally convergent, allow for a simple and unambiguous imposition of boundary conditions, and are very flexible to parallelization and adaptivity. These methods rely on imposing weak continuity of a numerical flux across the boundaries of the discretization elements, while solving the governing equations at the element level. The DG method was first applied to solve the time-domain Maxwell's equations by Hestaven and Warburton [101], and has been further developed to simulate wave propagation phenomena through metamaterials at the nanoscale [30], as well as for dispersive media [117,134,145] and more recently the hydrodynamic model [231] for 2d dimers.

Despite all these advantages, DG methods face limitations when used for practical 3d applications due to the computational burden that arises from nodal duplication. This shortcoming motivated the development of the hybridizable discontinuous Galerkin (HDG) method, first introduced in [48] for elliptic problems, subsequently analyzed in [47,50,52], and later extended to a wide variety of partial differential equations (PDEs) [49,169,173–177,255]. More specifically, the HDG has proven very effective for acoustics and elastodynamics [178, 225] and electromagnetics [179]. The HDG method is fully implicit, unstructured, and high-order accurate in both space and time, and is particularly effective for solving elliptic PDEs because it possesses several unique features that distinguish it from other DG methods. First, it reduces the number of globally coupled unknowns to those required to represent the trace of the approximate solution on the element boundaries, thereby resulting in a smaller global system than that of other DG methods [108]. Second, the method provides optimal convergence rates for both the solution and the flux. And third, its flux superconvergence properties can be exploited to devise a local postprocess that increases the convergence rate of the approximate solution by one order.

Traditionally, the main criticism shared by all approaches reviewed above is the treatment of unbounded problems. Modeling unbounded domains requires an artificial truncation of the computational domain in such a way that the waves propagating outwards do not reflect back and contaminate the solution. Common approaches are Sommerfeld radiation conditions, also known as absorbing boundary conditions (ABC) [67, 165, 242], applied at the mesh

boundaries where radiation is enforced, or perfectly matched layers (PMLs) [17, 95, 254]. PMLs are layers of absorbing material placed in the edges of the computational domain that exponentially attenuate outgoing waves, and are typically much more effective than ABCs.

1.3.2 Reduced order modeling

Proper orthogonal decomposition

The proper orthogonal decomposition (POD) has been extensively used due to its capacity to capture the essential information of an infinite-dimensional process with only a reduced number of basis functions. The POD consists of three equivalent methods, the Karhunen-Loève decomposition, the principal component analysis [122] and the singular value decomposition (SVD). We refer the reader to [72, 138, 146] and the references therein for a comprehensive discussion on the equivalence of these methods.

The most widely used approach to generate a low-dimensional representation of dynamical systems is the method of snapshots, originally introduced by Sirovich in [237]. POD requires a set of state solutions, usually called snapshots, obtained by numerically solving the full forward model for arbitrary values of the input parameters. The POD basis is then obtained by compressing the ensemble of snapshots using SVD techniques, such that with a reduced number of basis functions the dominant information of the system is retained. Furthermore, the elements of the POD basis are uncorrelated, and they constitute a low-rank representation of the state solutions, which is optimal among all low-rank approximations. The model is then obtained using a Galerkin projection onto the subspace spanned by this POD basis.

The POD method has been used for several large-scale dynamical systems, *e.g.* in CFD and aerodynamic applications [19, 97, 103, 270], optimal control of fluids [147, 211], turbomachinery flows [68, 271], microelectromechanical systems [109, 139] and electromagnetics [5, 230]. Furthemore, research has been devoted into extending POD for nonlinear systems and nonlinear structural dynamics [131, 158]. A thorough review of POD and its application to mechanical systems may be found in [127].

Other classes of relevant projection-based model order reduction techniques include Krylov

subspace methods [73,90] and balanced truncation [93,161].

Reduced Basis Method

The reduced basis (RB) method is a technique to obtain rapid yet accurate approximations of functional outputs of parametrized PDEs. It was first introduced in the early 1980s by Noor [183,184] for single and multiple parameter problems in nonlinear analysis of structures. Further work was developed to include *a priori* error analysis [78,204]. Recently, much work has been devoted to RB methods by Patera *et al.* [86–88, 148, 150, 151, 170, 220, 221, 257–260], introducing several new concepts that have greatly developed these techniques, such as: (1) the use of global approximation spaces based on snapshots of the solution for the full governing equations; (2) rigourous *a posteriori* error estimators to certify the quality of the approximation; and (3) the exploitation of an offline/online strategy to improve computational efficiency.

The first theoretical *a priori* convergence results by Maday *et al.* [151] demonstrated exponential convergence of the reduced-basis. The method was developed for linear elliptic problems with affine parametrization [149] and for eigenvalue problems [148]. Extensions to include nonlinear and noncoercive elliptic and parabolic problems were developed by Rovas, Veroy *et al.* [220, 256–260], together with developing rigorous and sharp error estimators.

However, sampling the parameter space is challenging, especially if high-dimensional inputs are considered. Recently, the greedy sampling method introduced by Patera *et al.* [86, 88, 257, 259] has proved to be an efficient strategy to adaptively select the snapshots based on estimates of the error. Alternatively, the greedy approach has also been reformulated as a sequence of adaptive model-constrained optimization problems [29], being advantageous from the point of view that the sample space is treated continuously.

In the context of multiscale and structured problems, a recent effort is the static condensation RB method [64, 111, 112], which exploits the structure by developing and reusing a RB for each different component, and then uses static condensation to recover a system on the interfaces that separate the components of the domain. Alternatively, the RB method has also been applied to the multiscale FE method [171], enabling an efficient approximation of the FE functions.

Recently, the RB method has been extended to the time-harmonic Maxwell's equations [42, 43, 55, 100] to successfully generate fast responses for complex electromagnetic devices, where both material properties and geometry are treated as parameters. Furthermore, the successive constraint method [110, 113] enables the computation of sharp and rigorous *a posteriori* error bounds.

The RB method has also been applied to standard continuous Galerkin finite element solutions of stochastic elliptic PDEs [25, 38, 94]. In this approach, the stochastic PDE is first reformulated as a parametrized PDE over the coefficients of the Karhunen-Loève expansion of the random fields. The reduced basis approximation and associated *a posteriori* error estimation are then developed for the resulting parametrized PDE. Finally, the output statistics and their error estimates are computed with a Monte Carlo simulation [25, 94] or a stochastic collocation approach [38]. These approaches, which involve the RB method and its *a posteriori* error bounds to evaluate the output instead of the original finite element discretization, have been shown to outperform standard stochastic simulation techniques.

An alternative approach for stochastic simulation are the recently introduced model and variance reduction (MVR) methods [261, 262], where the RB is combined with multilevel Monte Carlo sampling techniques [83]. This method enables the computation of statistical outputs and its *a posteriori* error estimates without involving *a posteriori* error bounds for the RB approximation, at a fraction of the cost of traditional MC methods. This feature broadens the applicability of RB methods to a wide variety of stochastic PDEs for which *a posteriori* error bounds for the RB approximation are either not available or too expensive to compute.

Empirical interpolation

The application of POD or RB techniques to linear PDEs or equations with an affine dependence on the relevant parameters constitutes a highly efficient approach for model reduction. However, the existence of nonlinearities in the problem poses a severe challenge for POD/RB, since the cost of evaluating the Galerkin projection scales as the dimension of the original system, hence barely improving the performance. For instance, we lose the attractive feature of precomputing the inner products needed to evaluate finite element formulations.

In order to overcome this limitation several alternatives have been proposed. The trajectory piecewise-linear method [216] proposes to generate a reduced model by combining linear models obtained linearizing along specific points on a state-space trajectory. The empirical interpolation method (EIM) [15] and its discrete counterpart DEIM [37] employ a greedy sampling strategy to generate an independent approximation to the nonlinear term, consisting of a weighted combination of interpolation functions. Thus, replacing the nonlinear term with this interpolation approximation recovers an efficient computational strategy when projecting onto the low-dimensional space. A similar methodology was developed in [172], where the greedy approach is replaced by an optimal selection of points based on a least-squares minimization, leading to the best points interpolation method. More recently, the localized DEIM [199] suggests the construction of multiple discrete interpolants for the nonlinear terms that correspond to different behaviors of the system.

1.4 Thesis objectives and overview

The main objective of this thesis is to propose, develop and validate a methodology for the efficient simulation of plasmonic devices. In particular, we aim to provide the community with a high fidelity numerical method and a reduced order modeling framework that enables the efficient simulation and design of metallic nanostructures. The accomplishments are summarized as follows:

- Implement an HDG method for the simulation of time-harmonic Maxwell's equations in 2d and 3d where the behavior of metals is characterized by the Drude model.
- Formulate, study and implement the HDG method for the hydrodynamic model for metals in 2d and 3d, thereby achieving a more realistic description of plasmonic resonances at the nanoscale.
- Formulate and implement a ROM for time-harmonic Maxwell's equations to account for parametric variability on the optical constants of metals and dielectrics, using proper orthogonal decomposition.
- Formulate and implement a ROM for time-harmonic Maxwell's equations for geometry modifications using a reference domain formulation, combining proper orthogonal

decomposition and empirical interpolation techniques.

- Validate the numerical simulations with experimental results reported in literature, and perform analyses to identify the sources of mismatch.
- Perform studies for the annular nanogap structure using the proposed methodology, in order to understand the effects of nonlocality and geometry on the transmission properties.
- Demonstrate the application of ROM to design a concentric annular nanogap structure with enhanced transmission capabilities.

In summary, we develop the HDG method for two hierarchical models to describe the electron motion in metals. The hydrodynamic model gives rise to a more physically accurate representation of the interaction of light with noble metals. This advanced model captures the nonlocal behavior of electrons that occurs at the nanometric scale by augmenting Maxwell's equations. In Chapter 2, we review the governing equations, physical models, and assumptions employed throughout the thesis. In Chapter 3, we formulate and implement the HDG method for the time-harmonic Maxwell's equations for both the LRA and the HM, and provide examples of realistic 3d devices that motivate the importance of considering the hydrodynamic model. In Chapter 4, we propose a ROM framework, based on proper orthogonal decomposition and empirical interpolation, to construct a surrogate model that can be evaluated in real-time, and yet provides accurate approximations to the high-fidelity model. In addition, the proposed method is capable of encoding parametric variation, for instance of optical constants or geometry parameters, which greatly facilitates simulation and design. In Chapter 5, we bring the above methodologies together and present several relevant applications in plasmonics, namely a saturation study for a 2d slit, a comprehensive parametric analysis of annular nanogaps, including detailed comparisons and validation with experimental data, and the design of a concentric annular structure. Finally, we conclude in Chapter 6 with some directions for future research.

Chapter 2

Governing equations and physical models

In this chapter, we review the basics of electromagnetic optics and introduce the governing equations that will be used throughout this thesis. The term electromagnetic optics stems from the treatment of light as an electromagnetic vector field, as opposed to wave optics, where light is modeled as a scalar wavefunction. The electromagnetic radiation is expressed as mutually coupled electric and magnetic field waves that propagate through the medium, and enables a more realistic treatment of the propagation and control of light. Indeed, describing electromagnetic radiation using vector fields allows us to account for light polarization, or a more comprehensive description of reflection and refraction at the interface between dielectric media. For the applications in this thesis, we shall focus on phenomena that occur at optical frequencies, which range from the far infrared to the ultraviolet (UV), that is from 300 GHz ($\lambda = 1$ mm) to 3 PHz ($\lambda = 100$ nm), as illustrated in Fig. 2-1.

The fundamental PDEs that describe electromagnetic wave propagation are the celebrated Maxwell's equations, which we review first, along with its constitutive relations. We devote special attention to the case of monochromatic waves, thus allowing to formulate the problem in frequency domain, along with suitable boundary conditions and relevant quantities of interest that will be used in the thesis. We then review the optical properties for both dielectrics and metals, and conclude by examining different physical models that have been proposed in literature to describe the conductivity of metals. These models are important to accurately describe the behavior of light at the interface between metals and dielectrics.

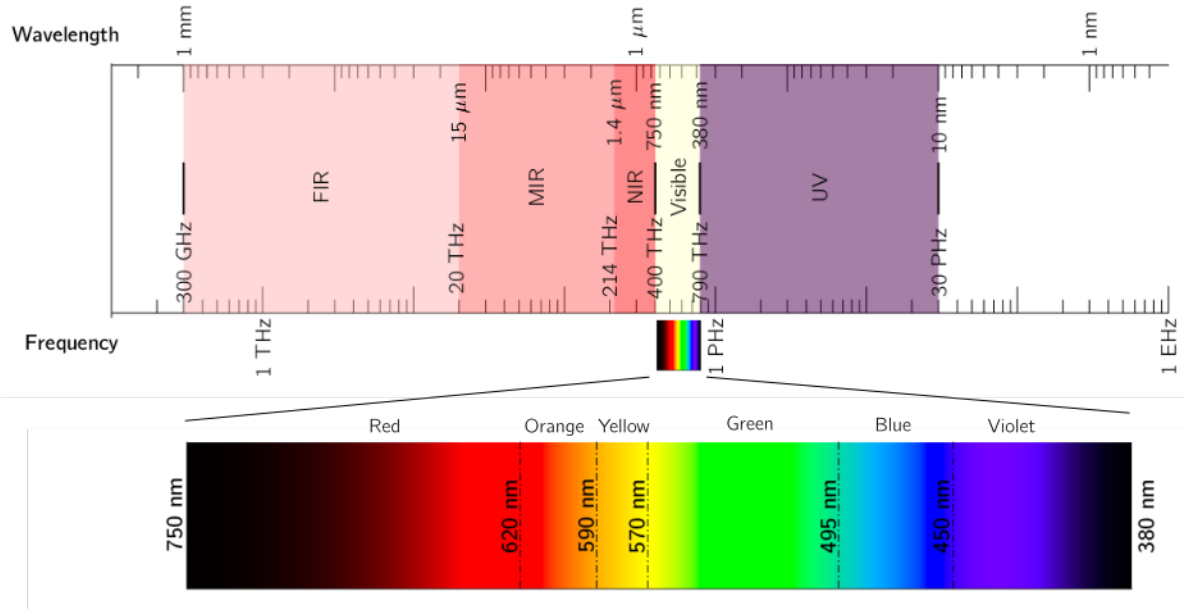


Figure 2-1: Optical frequencies and wavelengths. The infrared region (1 mm - 750 nm) of the spectrum is divided into far infrared (FIR), mid infrared (MIR) and near infrared (NIR). The visible comprises wavelengths from 750 nm to 380 nm, and the ultraviolet ranges from 380 nm to 10 nm.

.

2.1 Maxwell's Equations

The electric $\bar{\mathcal{E}}(\bar{\mathbf{x}}, \bar{t})$ and magnetic $\bar{\mathcal{H}}(\bar{\mathbf{x}}, \bar{t})$ fields are described by six scalar functions of space and time that completely characterize the propagation of light. These fields satisfy the following set of partial differential equations

$$\begin{aligned} \nabla \times \bar{\mathcal{E}} &= -\partial_{\bar{t}} \bar{\mathcal{B}} \quad \text{(Ampère's law)}, \\ \nabla \times \bar{\mathcal{H}} &= \bar{\mathcal{J}} + \partial_{\bar{t}} \bar{\mathcal{D}} \quad \text{(Faraday's law)}, \\ \nabla \cdot \bar{\mathcal{D}} &= \bar{\rho}, \quad \text{(Gauss's law)}, \\ \nabla \cdot \bar{\mathcal{B}} &= 0, \quad \text{(magnetic Gauss's law)}, \end{aligned} \tag{2.1}$$

known as Maxwell's equations. The quantities $\bar{\mathcal{D}}$, $\bar{\mathcal{B}}$ correspond to the electric displacement and magnetic flux density, respectively. In addition, we define the electric current and the charge density respectively as $\bar{\mathcal{J}}$ and $\bar{\rho}$. In addition to Maxwell's equations, we shall define

the following constitutive relations

$$\begin{aligned}
&\overline{\mathcal{B}} = \overline{\mu}\overline{\mathcal{H}}\,,\\
&\overline{\mathcal{D}} = \bar{\varepsilon}\overline{\mathcal{E}} = \varepsilon_0(1+\chi)\overline{\mathcal{E}} = \varepsilon_0\overline{\mathcal{E}} + \overline{\mathcal{P}}\,,\\
&\overline{\mathcal{J}} = \overline{\sigma}\overline{\mathcal{E}} + \overline{\mathcal{J}}^{im} \quad \text{(Ohm's law)},\\
&\bar{\varepsilon} = \varepsilon_0\varepsilon, \quad \overline{\mu} = \mu_0\mu.
\end{aligned} \tag{2.2}$$

The electric permittivity $\bar{\varepsilon}$ and the magentic permeability $\overline{\mu}$ are usually defined as a fraction of the free-space permittivity ε_0 and permeability μ_0, using the nondimensional relative permittivity ε and permeability μ, that is $\bar{\varepsilon} = \varepsilon_0\varepsilon$ and $\overline{\mu} = \mu_0\mu$. The free-space permittivity and permeability are related to the speed of light in vacuum as $c_0 = (\varepsilon_0\mu_0)^{-1/2}$. For an arbitrary medium, the permittivity and permeability are commonly represented as three dimensional positive definite tensors, but for isotropic media they reduce to scalars. The constant χ is the electric susceptibility, and $\overline{\mathcal{P}}$ represents the density of permanent or induced electric dipole moments in a dielectric material, usually referred to as polarization density, which vanishes in free space. For this work, we will assume linear (χ independent of $\bar{\mathcal{E}}$), non-magnetic ($\mu = 1$) and charge-free ($\rho = 0$) media. Finally, the quantity $\overline{\sigma}$ is the conductivity of the medium, which is zero for dielectric materials, and $\overline{\mathcal{J}}^{im}$ represents the impressed (or external) electric current.

For numerical stability, it is common to use a nondimensional version of Maxwell's equations. For a reference length scale L_c, we use the following scalings for the above fields

$$\begin{aligned}
\overline{\mathbf{x}} = \mathbf{x}/L_c, \quad \bar{t} = tc_0/L_c, \quad \bar{\mathcal{E}} = \alpha Z_0\mathcal{E}, \quad \overline{\mathcal{H}} = \alpha\mathcal{H},\\
\overline{\mathcal{D}} = \varepsilon_0\alpha Z_0\mathcal{D}, \quad \bar{\mathcal{B}} = \mu_0\alpha Z_0\mathcal{B}, \quad \overline{\mathcal{J}} = \alpha\mathcal{J}/L_c,
\end{aligned} \tag{2.3}$$

where α is a reference magnetic field and $Z_0 = \sqrt{\mu_0/\varepsilon_0}$ is the free-space impedance.

Applying this transformation to (2.1)-(2.2), we obtain

$$\begin{aligned}
\nabla\times\mathcal{E} &= -\partial_t\mathcal{B},\\
\nabla\times\mathcal{H} &= \mathcal{J} + \partial_t\mathcal{D},\\
\nabla\cdot\mathcal{D} &= 0,\\
\nabla\cdot\mathcal{B} &= 0,
\end{aligned}$$

and

$$\mathcal{D} = \varepsilon\mathcal{E} = (1+\chi)\mathcal{E} = \mathcal{E} + \mathcal{P}\,, \qquad \mathcal{B} = \mathcal{H}\,, \qquad \mathcal{J} = \sigma\mathcal{E} + \mathcal{J}^{im},$$

where σ is the nondimensional conductivity. Moreover, if we assume the electromagnetic waves are monochromatic, that is they propagate with a single frequency, further simplifications of the governing equations can be devised. For a certain (nondimensional) propagation frequency ν and angular frequency $\omega = 2\pi\nu$, we can write the components of, for instance, the electric field as $\mathcal{E}(\mathbf{x}, t) = \Re\{\mathbf{E}(\mathbf{x})\exp(-i\omega t)\}$, and analogously for the remaining fields. Consequently, time-dependence can be dropped, and Maxwell's equations are recast involving only the complex amplitudes. The resulting system is known as time-harmonic Maxwell's equations, or frequency domain formulation, and reads

$$\begin{aligned} \nabla \times \mathbf{E} - i\omega\mathbf{H} &= 0, \\ \nabla \times \mathbf{H} + i\omega\varepsilon\mathbf{E} &= \mathbf{J}, \\ \nabla \cdot \varepsilon\mathbf{E} &= 0, \\ \nabla \cdot \mathbf{H} &= 0. \end{aligned} \tag{2.4}$$

In order to complete the definition of the Maxwell's equations, we need to incorporate appropriate boundary conditions.

2.2 Boundary conditions

At the interface of two dielectric media the electromagnetic fields satisfy the following conditions

$$\mathbf{n}_1 \times \mathbf{E}_1 + \mathbf{n}_2 \times \mathbf{E}_2 = 0 \tag{2.5a}$$

$$\mathbf{n}_1 \times \mathbf{H}_1 + \mathbf{n}_2 \times \mathbf{H}_2 = 0 \tag{2.5b}$$

$$\mathbf{n}_1 \cdot \mathbf{B}_1 + \mathbf{n}_2 \cdot \mathbf{B}_2 = 0$$

$$\mathbf{n}_1 \cdot \mathbf{D}_1 + \mathbf{n}_2 \cdot \mathbf{D}_2 = 0$$

where $\mathbf{n}_1$, $\mathbf{n}_2$ are the unit outward normals of media 1 and 2 respectively, with $\mathbf{n}_1 = -\mathbf{n}_2$. We therefore have continuity of the tangential component of the electric and magnetic fields, continuity of the normal component of the magnetic flux density and a jump in the normal component of the electric flux density equal to the surface charge, here assumed zero. Con-

versely, at the interface between a dielectric and a perfect conductor ($\sigma = \infty$) the tangential component of the electric field vanishes, since inside a perfect conductor there is no electric field. Thus, the electric field has only normal component, that is the incident wave gets perfectly reflected (the conductor acts as a mirror). This boundary condition, known as perfect electric conductor (PEC), is expressed as $\mathbf{n} \times \mathbf{E} = 0$. Analogously, the condition $\mathbf{n} \times \mathbf{H} = 0$ is referred to as perfect magnetic conductor (PMC) boundary condition.

In addition, we should also consider the boundary conditions needed to enforce symmetry. This enables us to perform efficient numerical simulations by only accounting for the irreducible geometric structure, rather than the entire domain. We describe a monochromatic plane wave by its amplitude A_0, its propagation direction $\mathbf{d}$ and its polarization $\mathbf{p}$, such that $\mathbf{A} = A_0 \mathbf{p} \exp(i\omega\sqrt{\varepsilon\mu}\, \mathbf{d} \cdot \mathbf{x})$. For a 3d simulation with a plane wave propagating in the z-direction conditions must be prescribed in the edges of the transverse plane. For x-polarized waves, we have PEC for constant x and PMC for constant y. Conversely, if the incoming wave is y-polarized, we set PEC for constant y and PMC for constant x. This ensures the components of the electromagnetic field preserve symmetry specifications.

We finally discuss the numerical representation of unbounded domains. In order to simulate scattering from a volume, one must also take into account the medium surrounding the scatterer, which gives rise to a problem formulated in infinite domain. In finite element methods, the common strategy is to truncate the surrounding medium with an artificial surface Γ far from the scatterer enclosing the computational domain, and impose radiation conditions on Γ to recover a well-posed problem. The Silver-Müller radiation condition is a special case of ABCs [165, 242] for Maxwell's equations. It models the absorption of outgoing waves that propagate normally to the boundary Γ, preventing their reflection into the computational domain. This condition is first-order, since it approximates Γ with its tangent plane and involves only first derivatives of the fields, but satisfactory results are achieved if the scatterer is located far enough from Γ. If we decompose the electromagnetic fields into its incident and scattered parts, $\mathbf{E} = \mathbf{E}_0 + \mathbf{E}_s$ and $\mathbf{H} = \mathbf{H}_0 + \mathbf{H}_s$, the Silver-Müller condition is obtained by imposing

$$\mathbf{H}_s \times \mathbf{n} - Z^{-1} \mathbf{n} \times \mathbf{E}_s \times \mathbf{n} = 0,$$

where $\mathbf{n} \times \mathbf{E}_s \times \mathbf{n}$ is the tangent scattered electric field and $Z = \sqrt{\mu/\varepsilon}$ is the impedance at

the radiating boundary.

Another approach to simulate outgoing waves are perfectly matched layers (PMLs) [17], a numerical artifact that damps outgoing waves before they arrive to the boundary of the computational domain by appending additional layers that attenuate the physical solution. In order to use PMLs in numerical simulations, we transform the differential operators to include imaginary values through

$$\frac{\partial}{\partial x} \mapsto \frac{1}{1 - \frac{i\upsilon(x)}{\omega}} \frac{\partial}{\partial x},$$

where $\upsilon > 0$ turns the oscillating solution into an exponentially decaying one in the x-direction, whereas $\upsilon = 0$ leaves the equation unchanged. We refer to [120] for a detailed discussion of the PMLs and their implementation.

For the structures studied in thesis we have found the Silver-Müller conditions produce satisfying results, due to the extreme localization of the phenomena compared to the distance where radiation is imposed. Hence, the simulation examples presented are carried out with first order ABCs, less computationally demanding than PMLs.

2.3 Intensity and power

The flow of electromagnetic power is given by the Poynting vector, defined as $\mathcal{S} = \mathcal{E} \times \mathcal{H}$. In frequency domain, the Poynting vector is computed by time-averaging $\mathcal{S}$ over times that are longer than an optical cycle, thus invoking the time-harmonic nature of the electromagnetic fields we arrive at the following expression for the electromagnetic power $\mathbf{S}$

$$\mathbf{S} = \frac{1}{2} \Re \left[\mathbf{E} \times \mathbf{H}^* \right], \tag{2.6}$$

where $*$ refers to the complex conjugate.

For the case of plane wave scattering, it is important to describe the various versions of electromagnetic power that we encounter [21]. Consider a scatterer embedded in air and illuminated by a plane wave. The power generated by the scatterer is computed by integrating (2.6) on an arbitrary surface A surrounding the scatterer, hence the power absorbed

is defined as the negative of the generated power, $P_{abs} = -\int_A \mathbf{S} \cdot d\mathbf{A}$. Since we assume the surrounding medium is lossless, all the absorption is undertaken by the scatterer, and the integral can be performed on any enclosing surface. Rewriting the electromagnetic fields into the incident and scattered parts, the latter integral can be decomposed as

$$-P_{abs} = \frac{1}{2}\int_A \Re\left[\mathbf{E}_0 \times \mathbf{H}_0^*\right] \cdot d\mathbf{A} + \frac{1}{2}\int_A \Re\left[\mathbf{E}_s \times \mathbf{H}_s^*\right] \cdot d\mathbf{A} + \frac{1}{2}\int_A \Re\left[\mathbf{E}_0 \times \mathbf{H}_s^* + \mathbf{E}_s \times \mathbf{H}_0^*\right] \cdot d\mathbf{A}.$$

The first integral corresponds to the incident power, which vanishes as there are no losses in the medium. The second integral is power scattered by the object P_{sca}, whereas the last term is commonly known as the negative of the extincted power $-P_{ext}$. Rearranging the terms we arrive to $P_{ext} = P_{sca} + P_{abs}$, that is the extincted power represents the total radiant power scattered and absorbed by the object.

2.3.1 Quantities of interest

When numerically simulating the interaction of an electromagnetic field with heterogeneous materials, it is common to study its response by computing one or several figures of merit or quantities of interest (QoI). These QoI are typically obtained by evaluating a numerical expression that involves the solution of the full electromagnetic simulation. The QoI that will be used throughout this dissertation are the optical intensity, the field enhancement and the extinction cross section.

The optical intensity reflects the amount of electromagnetic power that flows through a surface A with normal vector $\mathbf{n}$, and is defined as

$$\langle \mathbf{S} \rangle = \frac{1}{2}\int_A \left|\Re\left[\mathbf{E} \times \mathbf{H}^*\right] \cdot \mathbf{n}\right| dA\,.$$

Alternatively, performance of electromagnetic devices is usually analyzed by measuring the transmitted power, or transmitted optical intensity. The transmitted power ς can be evaluated as

$$\varsigma = \frac{\langle \mathbf{S} \rangle}{\langle \mathbf{S}_0 \rangle}$$

where $\langle \mathbf{S}_0 \rangle$ refers to the optical intensity of the incident electromagnetic field.

The field enhancement measures the local amplification factor of the incoming electromag-

netic field that occurs in a tightly confined region V of the domain. The field enhancement will be measured as the spatial average of the ratio of electric amplitudes, that is

$$\pi = \frac{1}{|V|} \int_V \left| \frac{\mathbf{E} \cdot \mathbf{p}}{\mathbf{E}_0 \cdot \mathbf{p}} \right| dV \,,$$

where $\mathbf{p}$ refers to the direction of polarization of the incident field.

Finally, we define the extinction cross section σ_{ext}, that is computed by scaling the extincted power by the cross section σ of the scatterer and the incident intensity, which for plane wave illumation reads $|\mathbf{E}_0|^2/2$, namely

$$\sigma_{ext} = 2 \frac{P_{ext}}{\sigma |\mathbf{E}_0|^2}.$$

The extinction cross section is also a dimensionless quantity, and represents the absorption and scattering strength of an object at a given wavelength.

2.4 Optical properties of materials

In order to control and manipulate light we employ metamaterials, which consist of a combination of homogeneous materials usually arranged in lattice structures at the microscopic level. The remarkable property of metamaterials is achieved by leveraging the optical interaction of its constituent materials and the periodic arrangement to attain phenomena that cannot be observed for homogeneous materials. In this section, we review the main features of dielectrics and metals, which are commonly used as elemental components for metamaterials.

The main physical difference between metals and dielectrics are the energy bands, that is regions of energy where electrons are allowed. In a simplified version, we have the valence band, which is fully occupied by electrons in the outermost orbit that are bound to the ion cores, and the conduction band, which represents higher energy levels that are mostly empty. The electrons in the conduction band are the ones involved in the conduction process, since they are not attached to any atom. These two bands are separated by an energy band where electrons are disallowed, usually known as band gap. Hence, for an electron in the valence band to reach the conduction band it must overcome the energy barrier posed by the band

gap.

Dielectric materials do not exhibit flow of electric charges under an electric field. The main cause of their negligible conductivity is the fact that they possess large band gaps, thus an enormous amount of energy needs to be supplied to the electrons in the valence band to reach the conduction band. Consequently, dielectrics act as insulators in the presence of an electric field. The permittivity ε of the dielectric is closely related to the susceptibility χ, such that low susceptibility values imply higher resistance to be polarized in the presence of an electric field. For the majority of applications in this thesis we will consider that both ε and the refractive index $n = \sqrt{\varepsilon}$ are real quantities, that is dielectrics behave as lossless materials.

Conversely, the band gap in metals is nonexistent, hence a minimum supply of energy drives the electrons in the valence band to the conduction band to participate in the conduction process. Metals therefore behave as conductors when an electric field is applied, since there are always electrons available in the conduction band. Even though ideal conductors have an infinite conductivity, for realistic applications the characterization of their frequency-dependent permittivity and conductivity is still an active field of research. Below we review the main efforts to model the optical properties of metals.

2.5 Modeling light-metal interaction

In this section, we summarize and review the most relevant approaches used to model the interaction of metals with light, devoting special attention to the nanometer and subnanometer regimes. We first present the two main classical models that will be implemented in this dissertation, and analyze their differences for nanoparticle systems. Finally, we survey alternative and complementary models that have been proposed in literature to represent quantum effects. Even though these more complicated avenues will not be further investigated in this thesis, they are introduced with the intention to provide a complete view of the state-of-the-art approaches, and left as future lines of research.

2.5.1 Local response approximation

In order to effectively simulate electromagnetic wave propagation through metals we have to take into account their dispersive nature. A classical and widely-used model for the permittivity of a metal is the Drude model [61], which assumes the electrons in the valence band are fully detached from the ions, and thus behave as particles of an ideal gas. This model is commonly known in literature as local model or local response approximation (LRA) for metals, since the motion of an electron is not coupled to the rest. In order to derive the Drude model we use the time-dependent Maxwell's equations in (2.1), and write the equation of motion for a single electron in the absence of impressed currents. The equation is given by a mass oscillator under an electric field with damping, that is

$$-e\overline{\mathcal{E}} = m_e \partial_{\bar{t}\bar{t}} \overline{\mathbf{x}} + m_e \overline{\gamma} \partial_{\bar{t}} \overline{\mathbf{x}}$$

where m_e is the effective electron mass, e is the charge of the electron and $\overline{\gamma}$ is a damping constant related to the collision rate of the electrons. The polarization $\overline{\mathcal{P}}$ can be split into a frequency-independent contribution $\overline{\mathcal{P}}_\infty$ that depends on the bound electrons in the valence band and a polarization arising from the freely moving electrons in the conduction band $\overline{\mathcal{P}}_f$. This latter polarization is related to the displacement of an electron $\overline{\mathbf{x}}(\bar{t})$ as $\overline{\mathcal{P}}_f = -ne\overline{\mathbf{x}}(\bar{t})$ for a given electron density $n(\overline{\mathbf{x}}, \bar{t})$, and to the current density $\overline{\mathcal{J}}$ as $\overline{\mathcal{J}} = \partial_{\bar{t}} \overline{\mathcal{P}}_f$. Substituting this relationship into the equation for the motion of an electron we arrive to

$$\frac{e^2 n}{m_e \varepsilon_0} \varepsilon_0 \overline{\mathcal{E}} = \partial_{\bar{t}\bar{t}} \overline{\mathcal{P}}_f + \overline{\gamma} \partial_{\bar{t}} \overline{\mathcal{P}}_f.$$

The parameter involved in the first term is the square of the metal's plasma frequency $\overline{\omega}_p$, defined as $\overline{\omega}_p^2 = \dfrac{e^2 n}{m_e \varepsilon_0}$. The plasma frequency represents the frequency above which the conduction electrons are not able to oscillate in phase with the incident light, thus effectively impeding the cancellation the incoming wave. That is, for $\omega > \omega_p$ metal behaves as a dielectric, and the incident wave is allowed to propagate through the metal (with losses). However, for all the applications in this thesis we consider the regime $\omega < \omega_p$.

We now invoke the scalings in (2.3), rendering the nondimensional plasma frequency $\omega_p = \overline{\omega}_p L_c / c_0$ and collision rate $\gamma = \overline{\gamma} L_c / c_0$, and apply the frequency domain transformation $\partial_t \mapsto -i\omega$ to compute the dispersive free-electron polarization as $\mathbf{P}_f = \frac{i}{\omega}\mathbf{J} = -\frac{\omega_p^2}{\omega(\omega + i\gamma)}\mathbf{E}$.

Thus, by Ohm's law the conductivity may be expressed as $\sigma = i\omega_p^2/(\omega + i\gamma)$, and if we combine it with Faraday's law, we arrive at a complex dispersive permittivity for metals, also known as Drude model, namely

$$\varepsilon_D(\omega) = \varepsilon_\infty - \frac{\omega_p^2}{\omega(\omega + i\gamma)}, \tag{2.7}$$

where ε_∞ is the frequency-independent core electron permittivity that arises from the polarization $\mathbf{P}_\infty$, that is $\mathbf{E} + \mathbf{P}_\infty = (1 + \chi_\infty)\mathbf{E} = \varepsilon_\infty \mathbf{E}$. The imaginary part of (2.7) is typically very large, incurring a considerable loss or dissipation that minimizes the propagation of electromagnetic waves in the metal. In addition, the real part of (2.7) exhibits negative values, hence the free electrons in the metal are out of phase with respect to the incident electric field. Consequently, metals act as a mirror, that is the majority of incident light is reflected.

The optical constants have been extensively studied in order to determine their values for different frequency regimes. Although we shall specify the values used for all the examples provided in this dissertation, we refer the reader to the classical references [119,192,193,208] for the experimental determination of the optical constants for metals.

The Drude model for metals is attractive for its simplicity, and produces satisfying results for many electromagnetic applications. Nonetheless, the assumption that all electrons are free electrons and have a local behavior is rather simplistic, and may be inaccurate for frequencies close to the plasma frequency. An extension to the Drude model is the Drude-Lorentz model, which accounts for interband transitions –electrons breaking its covalent bond with the ion and transitioning from the valence band to the conduction band– that are relevant at high frequencies. The idea behind this approach is to model the interaction between the free electrons and the ion cores as an oscillator, which results in augmenting (2.7) with Lorentzian peaks. This extension is beyond the scope of this thesis, since for the applications of interest the interband transitions can be neglected. We refer the reader to [32] for further details.

Furthermore, the Drude model has been shown to produce unphysical solutions when predicting light-metal interaction at geometries below tenths of nanometers [219]. A more realistic model for noble metals may be obtained by assuming the free electrons do not be-

have independently, rather their motion is nonlocally coupled. This effect can be accounted for with the hydrodynamic model, which is described next.

2.5.2 Hydrodynamic model

A hydrodynamic model for the free electron gas was introduced in the 1970s [65]. This model, despite neglecting quantum phenomena such as quantum tunnelling and quantum oscillations, introduces a hydrodynamic pressure term that accounts for the inter-electron coupling, or nonlocal interaction, that becomes relevant for geometries below tenths of nanometers. Hence, it is referred to as nonlocal model or hydrodynamic model (HM) for noble metals. The HM, which has recently been studied from the finite element perspective [102,231], has been shown to produce better experimental agreement for nanoparticles and dimers [157,212,250] than the LRA.

A thorough derivation of the HM can be found in the literature [20, 65, 203]. In this thesis we merely review the most important aspects of the derivation to gain physical insight. We introduce the electron density $n(\overline{\mathbf{x}}, \bar{t})$, the electron pressure $p(\overline{\mathbf{x}}, \bar{t})$ and the hydrodynamic velocity $\mathbf{v}(\overline{\mathbf{x}}, \bar{t})$, which are related by the continuity equation as $\partial_{\bar{t}} n = -\nabla \cdot (n\mathbf{v})$. The equation of motion for the electron fluid under a macroscopic electromagnetic field is described as

$$m_e(\partial_{\bar{t}} + \mathbf{v} \cdot \nabla + \overline{\gamma})\mathbf{v} = e(\overline{\mathcal{E}} + \mathbf{v} \times \overline{\mathcal{H}}) - \frac{\nabla p}{n}. \tag{2.8}$$

In order to reduce the above equation, some simplifications need to be made. We linearize the electron density field around the constant equilibrium density of the electron gas n_0 such that $n(\overline{\mathbf{x}}, \bar{t}) \approx n_0 + n_1(\overline{\mathbf{x}}, \bar{t})$, neglect the high order term for the derivative of the hydrodynamic velocity $\mathbf{v} \cdot \nabla \mathbf{v}$ and also neglect the effect of the magnetic field, since the electron fluid is driven mainly by the electric field. In addition, we simplify the pressure term in (2.8) assuming a Thomas-Fermi model where only the kinetic energy is relevant, that is

$$\frac{\nabla p}{n} \approx m_e \overline{\beta}^2 \frac{\nabla n_1}{n_0}.$$

The quantum parameter $\overline{\beta}$, which represents the nonlocality, is usually expressed [20] in terms of the Fermi velocity $\overline{v}_F$ of the metal as $\overline{\beta} = \sqrt{3/5}\overline{v}_F$. Using the assumptions above,

the equation of motion for the electron fluid can be further simplified as

$$m_e(\partial_{\bar{t}} + \overline{\gamma})\mathbf{v} = -e\overline{\mathcal{E}} - m_e\overline{\beta}^2\frac{\nabla n_1}{n_0},$$

and if we differentiate with respect to time, we arrive at

$$m_e(\partial_{\bar{t}\bar{t}} + \overline{\gamma}\partial_{\bar{t}})\mathbf{v} = -e\partial_{\bar{t}}\overline{\mathcal{E}} + m_e\overline{\beta}^2\nabla(\nabla\cdot\mathbf{v}), \tag{2.9}$$

where the last term is obtained by linearizing the continuity equation $\partial_{\bar{t}}n_1 = -\nabla\cdot(n_0\mathbf{v})$ and neglecting the high-order term $\nabla\cdot(n_1\mathbf{v})$. Using the relation between the electric current and the electron gas density $\overline{\mathcal{J}} = -en\mathbf{v}$, and multipliying (2.9) by $-en/m_e$, we obtain

$$\partial_{\bar{t}\bar{t}}\overline{\mathcal{J}} + \overline{\gamma}\partial_{\bar{t}}\overline{\mathcal{J}} = \frac{e^2n}{m_e}\partial_{\bar{t}}\overline{\mathcal{E}} + \overline{\beta}^2\nabla(\nabla\cdot\overline{\mathcal{J}}).$$

If we invoke again the scalings in (2.3) and express the equation in frequency domain, we arrive at

$$\beta^2\nabla(\nabla\cdot\mathbf{J}) + \omega(\omega + i\gamma)\mathbf{J} = i\omega\omega_p^2\mathbf{E}, \tag{2.10}$$

for $\beta = \overline{\beta}/c_0$. The result in (2.10) imposes a nonlocal relation between the electric current and the electric field, thus if we pursue numerical simulations in a metal we have to augment Maxwell's equations (2.4) with (2.10) and solve the coupled system of equations, namely

$$\begin{aligned}
\nabla\times\mathbf{E} - i\omega\mathbf{H} &= 0,\\
\nabla\times\mathbf{H} + i\omega\varepsilon_\infty\mathbf{E} &= \mathbf{J},\\
\beta^2\nabla(\nabla\cdot\mathbf{J}) + \omega(\omega + i\gamma)\mathbf{J} &= i\omega\omega_p^2\mathbf{E},\\
\nabla\cdot\varepsilon_\infty\mathbf{E} &= 0,\\
\nabla\cdot\mathbf{H} &= 0.
\end{aligned} \tag{2.11}$$

The hydrodynamic model above will be used for the examples throughout the thesis.

A more general model for the study of plasmon excitation has recently been introduced [164], and accounts not only for the electron pressure, but also for the diffusion kinetics produced by the induced charges. To that effect, the proposed generalized nonlocal optical response (GNOR) model is effectively a convection-diffusion equation for the hydrodynamic current.

The effect of the GNOR model combines the usual blueshifts due to the electron pressure with a degradation on the plasmonic resonant peak caused by the diffusion.

Mathematically, the diffusion can simply be encoded in the quantum parameter β through the modification $\beta^2 \mapsto \beta^2 + D(\gamma - i\omega)$, where D is a (nondimensional) diffusion constant. The main difference with respect to the pure convection case is that the quantum parameter has now an imaginary contribution, proportional to the frequency, which produces a damping effect. The value of the diffusion parameter depends on the probability of electrons scattering off the surface of the particle. Its experimental determination, as well as further discussions on the implications of GNOR, may be found in [164, 214].

Despite the enhanced predictive power of HM compared to LRA, there are situations in which it still fails to capture the behavior observed experimentally. For instance, the HM is unable to correctly describe the resonant frequency red-shifts for alkali metals that have been reported both experimentally [215] and numerically [136], instead of the blue-shifts observed for noble metals, that are correctly predicted by the HM. In addition, it neglects quantum phenomena such as electron tunnelling, that become relevant for subnanometric geometries. Hence, the HM is appropriate for plasmonic structures with noble metals in the nanometer and supernanometer scales [79, 279].

2.5.3 Comparison between local response approximation and hydrodynamic model

In this section, we compare the HM to the LRA in order to better understand the distinctions between both approaches from the physics and computational perspective.

Single cylindrical nanowire

In order to demonstrate the HM, we consider a golden nanowire of diameter D in free space. We assume the nanowire is infinite in the z direction, and is excited by a x-polarized electric field propagating the y-direction, that is $\mathbf{E}_0 = E_0(\exp(i\omega y), 0, 0)$, see Fig. 2-2. For this simple geometry the analytical solution is available for both the LRA and the HM model using Bessel and Hankel functions [222], and is useful to illustrate the physics captured by both models. In Fig. 2-3, we depict the extinction cross section σ_{ext}, for both the local

($\beta = 0$) and the nonlocal model ($\beta > 0$), for diameters 4 nm and 40 nm. The values for the gold constants are $\varepsilon_\infty = 1$, $\hbar\overline{\omega}_p = 9.02$ eV and $\hbar\overline{\gamma} = 0.071$ eV, taken from [119].

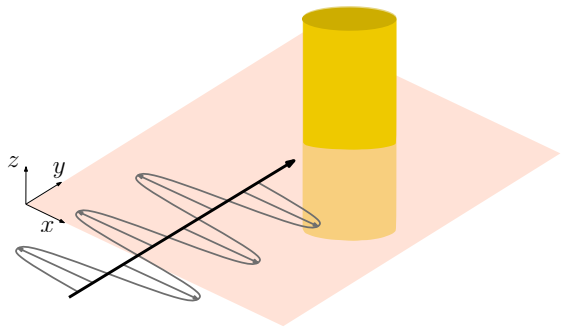
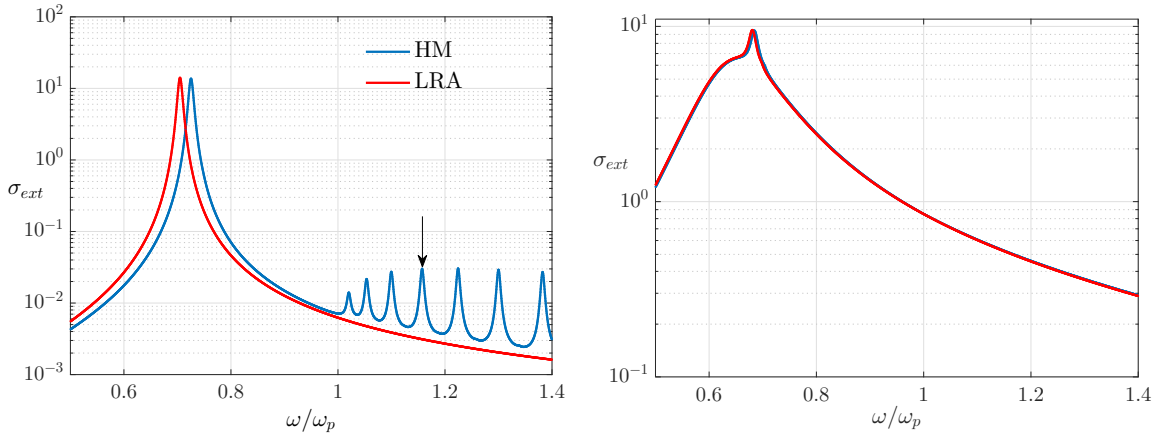


Figure 2-2: Schematic diagram of single nanowire illuminated with electromagnetic wave.

As anticipated, for small metallic nanoparticles the effects of the hydrodynamic current are significant, causing not only a blueshift of around 3% in the main resonance, but also a sequence of resonances for frequencies above the plasma frequency that are not excited with the local model. Conversely, the effects of the HM are virtually nonexistent for the 40 nm wire, and the curves for σ_{ext} are almost overlapping. Therefore, the HM is able to capture excitation of plasmons that are omitted by the LRA, such as the longitudinal modes explained in the next section.



(a) $D = 4$ nm (b) $D = 40$ nm

Figure 2-3: Extinction cross section of gold nanowire in air for LRA and HM.

Excitation of longitudinal modes

For a plane wave polarized in a direction $\mathbf{p}$ transverse to the propagation direction $\mathbf{d}$, its dispersion relation is given by $k^2 = \varepsilon_D(\omega)\omega^2$, for the dispersive Drude permittivity (2.7). In order to excite longitudinal modes ($\mathbf{k} \parallel \mathbf{E}$) we need $\varepsilon_D(\omega) = 0$, which for the extreme case $\gamma \to 0$ only happens at the plasma frequency. However, with the inclusion of the electron pressure term a continuum of longitudinal modes are supported, since the permittivity now depends on the propagation vector. A simple manipulation of (2.10), assuming $\mathbf{E}$ and $\mathbf{P}_f$ are parallel to $\mathbf{k}$, casts the longitudinal dielectric function of the metal

$$\varepsilon_L(\omega) = \varepsilon_\infty - \frac{\omega_p^2}{\omega(\omega + i\gamma) - \beta^2 k^2}.$$

The modes that propagate in the longitudinal direction satisfy $\varepsilon_L = 0$, and correspond to oscillations of the free electron gas due to the hydrodynamic current, also known as bulk plasmons.

The quantum parameter β controls the level of nonlocality, since as $\beta \to 0$ we recover the local Drude model. Conversely, for nonzero β the excited longitudinal modes exhibit a dispersive behavior. It can be shown [46, 163] that below the plasma frequency both the transverse and the longitudinal modes decay exponentially, whereas above the plasma frequency both modes propagate. As a matter of fact, it is the propagation of the longitudinal modes that causes the resonances shown in Fig. 2-3a for $\omega > \omega_p$. Since LRA only allows a longitudinal mode at the plasma frequency, these additional resonances are not excited in the local model.

Electron density

A compelling difference between LRA and HM is the distribution of the induced free charge density ρ_f, which is defined as $\rho_f = \nabla \cdot \mathbf{P}_f = -i\omega\nabla \cdot \mathbf{J}$. The solutions provided by the local model infinitely squash ρ_f at the metal surface, which results in a Dirac delta at the metal-dielectric interface. That is, the metal acts as a hard wall for the incoming EM wave, and impedes propagation through it. Conversely, the electron pressure term in the hydrodynamic model regularizes the induced charge density by smoothing its profile,

thus allowing the penetration of the incident field. The spreading distance experienced by the charge density is controlled by the δ parameter introduced by Ciracì *et al.* [46], that essentially measures the amount of nonlocality as $\delta = \beta/\omega_p$.

The impact of the hydrodynamic current in the charge density is significant, since the resonance shifts predicted by the HM below tenths of nanometers are explained by the effective diffusion of the metal-dielectric boundaries.

2.5.4 Simulation of hydrodynamic model

The inclusion of the hydrodynamic current poses notable challenges from the numerical simulation standpoint. Firstly, we incorporate an additional second-order equation, thus the degrees of freedom for the electric current need to be simultaneously solved. This circumstance impacts both the computational power and storage requirements. The boundary conditions introduced before do not suffice to uniquely determine the response, thus additional boundary conditions are needed at the metal interface. As suggested in literature [163], the appropriate boundary condition is $\mathbf{n} \cdot \mathbf{J} = 0$ at the interface of the metal. This condition precludes the electrons from leaving the metal (no electron spill-out), hence the normal electric current must vanish at the interface.

Secondly, the inclusion of the electron pressure term excites features that occur at the sub-Fermi-wavelength scale. This wavelength is associated to the Fermi energy –the maximum energy of the electrons in the metal– and is typically much smaller than the length scale of the problem. It is therefore necessary to perform the simulations on finer spatial discretizations, thus making its application to realistic 3d problems more difficult. In order to illustrate this phenomenon, we revisit the 4 nm wire in Fig. 2-2, and show $|\mathbf{E}_y|$ for both the local and the nonlocal model in Fig. 2-4. Results are computed with the HDG method introduced in Chapter 3 for the resonant frequency $\omega/\omega_p = 1.157$ indicated with an arrow in Fig. 2-3. Even though the solution outside the metal is similar, the hydrodynamic current excites resonances in the interior of the conductor at frequencies higher than the plasma frequency, resulting in ripples inside the nanowire of wavelength 100 times smaller than the wavelength of the incident field, see Figs. 2-4b and 2-4c. Consequently, to properly capture the nonlocal effects we require significantly finer discretizations within the metal. Finally, the incidence of

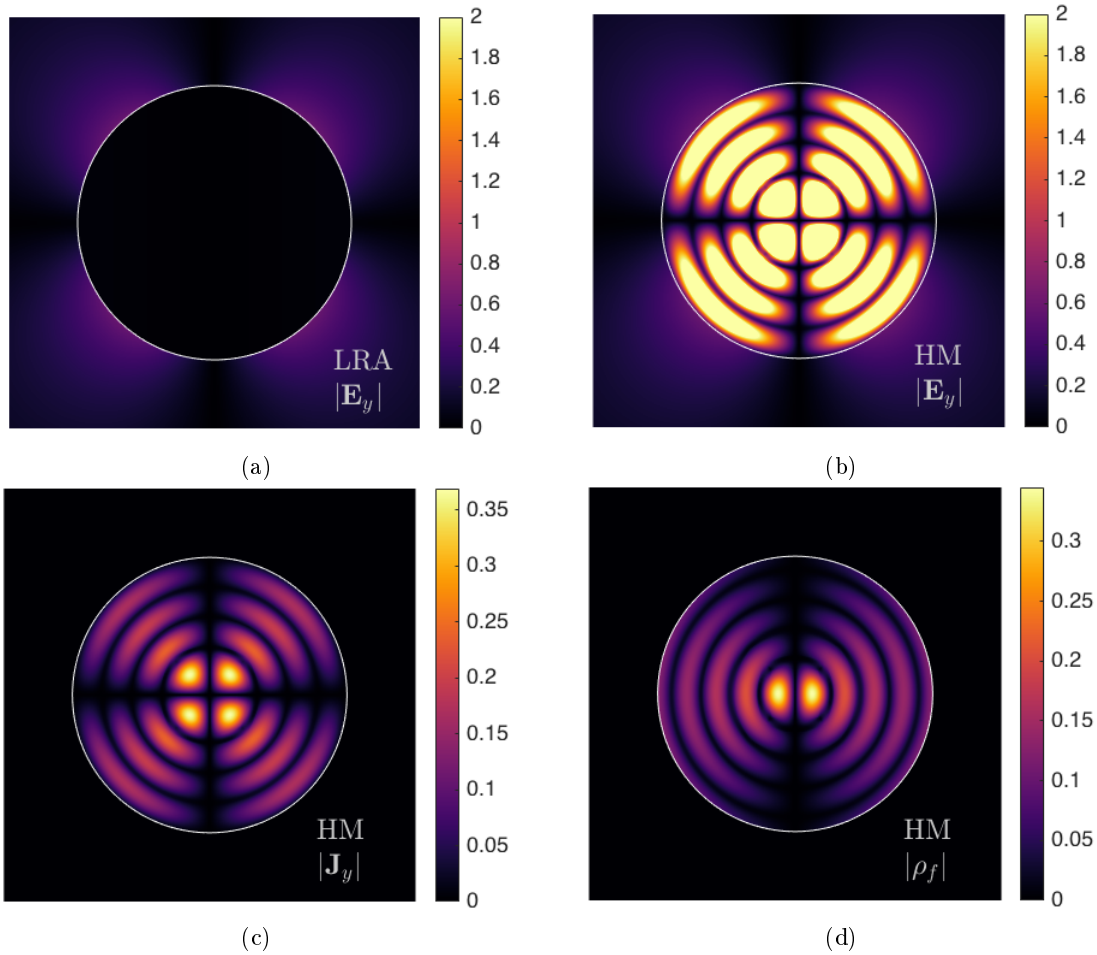


Figure 2-4: Solution fields for 4 nm gold wire at $\omega/\omega_p = 1.157$, with boundary highlighted.

the electron pressure is noticed by inspecting the small, albeit nonzero, induced free charge density in the metal, shown in Fig. 2-4d, as predicted by the hydrodynamic model.

In light of these circumstances, it is of the utmost importance to devise high-fidelity simulation methods that enable an efficient treatment of multiple scales, along with non-uniform curved meshing tools to accurately model the intricate geometries. In addition, we require a framework that is amenable to multiphysics simulations, since metals and dielectrics now have different governing equations, and thus require suitable compatibility conditions at the interface. This compatibility conditions between a dielectric medium 1 and a metal 2 are continuity of the tangential component of the electromagnetic fields (2.5a)-(2.5b) and zero jump in the normal component of the current $\mathbf{n}_1 \cdot \mathbf{J}_1 + \mathbf{n}_2 \cdot \mathbf{J}_2 = 0$, which reduces to the above condition since $\mathbf{J}_1 = 0$ in the dielectric by definition.

2.5.5 Application to periodic annular nanogap

We now consider a 3d structure, the annular gap, consisting of a periodic array of ring-shaped apertures patterned in a metal film. In order to focus only on the impact of the metal, we shall assume the film is suspended in free space (no substrate), and that there is no material filling the nanometer-wide gap. Although this structure cannot be manufactured, it will be of great interest to achieve a deeper understanding of the ring structure from a theoretical perspective.

The periodic structure is sketched in Fig. 2-5a, along with the illumination and dimensions used. The results presented in this section are obtained using the high-fidelity simulation techniques introduced in Chapters 3 and 4, thus we defer the technical details of the numerical simulations, and focus only on the physical implications of the results. In order to exhaustively study the structure among different regimes, we consider aperture widths w ranging from 5 Å to 100 nm, for frequencies between 0.2 and 5.5 THz, and investigate the response using the distinct models for light-metal interaction introduced above.

The simplest model is assuming the gold film behaves as a perfect conductor with infinite conductivity. Prescribing PEC conditions at the metal interface ensures the electric field is reflected at the metal boundary and no penetration is allowed. The field enhancement profile is shown in Fig. 2-5b with solid lines, exhibiting sharp peaks and enormous enhancements across gapsizes, showing that smaller gaps lead to larger resonances. This response corresponds to that of an undamped oscillator, which differs significantly to what has been observed experimentally for arrays of annular nanogaps [14]. Quite interestingly, this unrealistic behavior may also be observed with the undamped Drude model ($\gamma = 0$). The field enhancement curves for this case, using $\hbar\overline{\omega}_p = 9.02$ eV and $\varepsilon_\infty = 1$ adopted from Ordal *et al.* [192, 193], are also depicted in Fig. 2-5b with dashed lines. We note that the maximum enhancement attained with undamped Drude and with PEC models is identical for a given gap size. Hence, the collision rate plays a pivotal role in the accurate characterization of the electromagnetic response through Drude's permittivity, since it is responsible for the imaginary component that models losses in the metals.

Secondly, we introduce damping in the Drude model, with $\hbar\overline{\gamma} = 0.02678$ eV given by [192, 193], otherwise known as LRA. In Fig. 2-5c, we collect the field enhancement profile

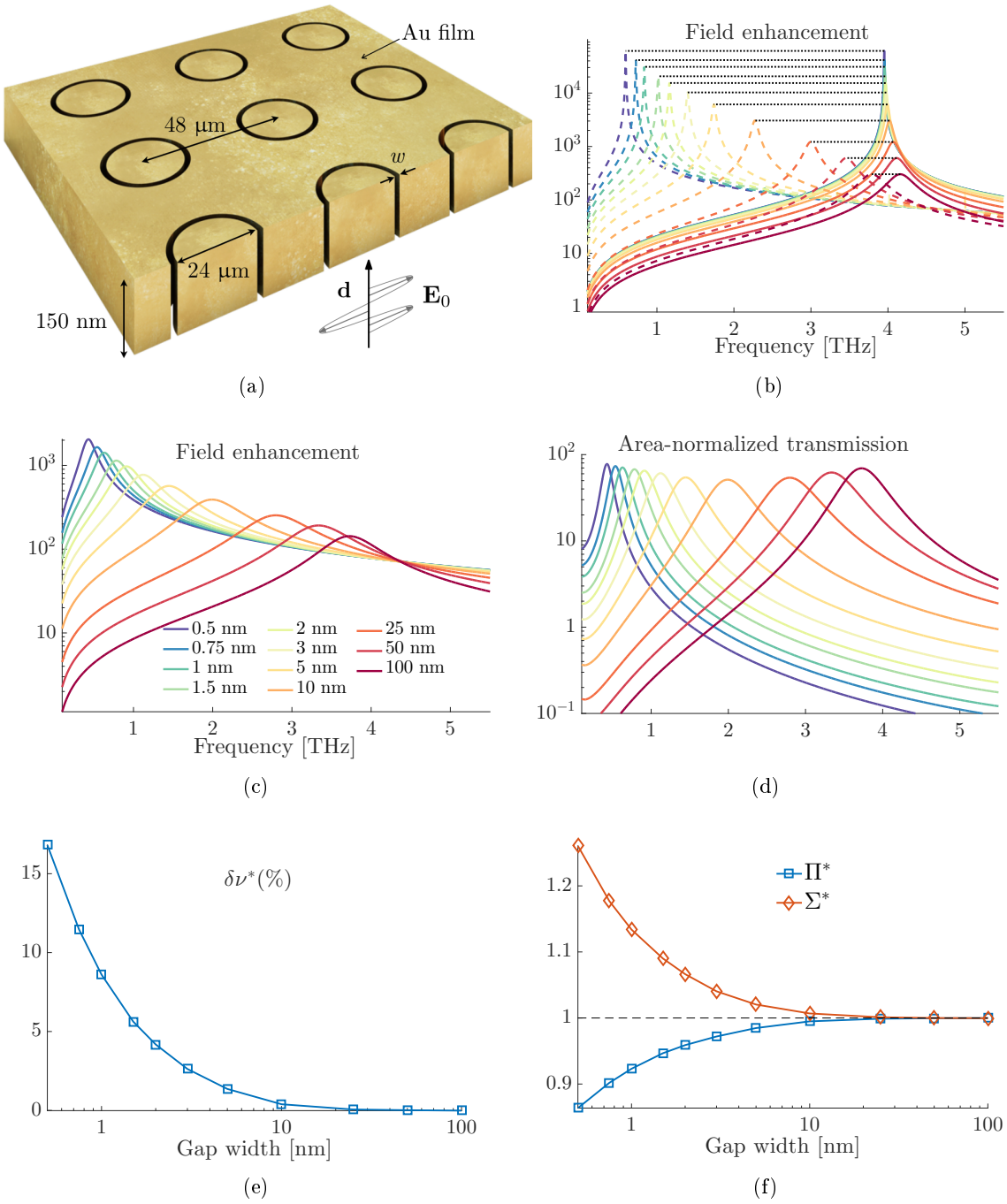


Figure 2-5: (a) Schematic of periodic array of annular gaps with relevant dimensions. (b) Field enhancement for PEC (solid) and undamped Drude (dashed). (c) Field enhancement for damped Drude. (d) Area normalized transmission for damped Drude. (e) Relative blueshift introduced by nonlocality. (f) Impact of nonlocality in field enhancement and transmission.

for various nanogaps. The difference between damped Drude and both PEC and undamped Drude are mainly lower enhancements and broader resonances, as a consequence of the losses introduced by a nonzero damping. Among distinct gap widths, these profiles are qualita-

tively similar, although smaller apertures lead to stronger field localizations and narrower resonance peaks. In addition, we present the transmission power of these structures in Fig. 2-5d, normalized by the open area fraction, that is $A_w/(A_w + A_{\mathrm{Au}})$. The metal is a lossy medium, thus the immense majority of light transmitted by the gold film is through the gap. Normalization ensures smaller widths are balanced by the superior excitation caused by plasmonic resonances. Indeed, the normalized transmission for nanometric and subnanometric gaps is superior to that of nanogaps 100 times wider, as a consequence of the extreme amplification of the incident EM field that occurs for deep-subwavelength apertures.

Finally, we extend the study above with the hydrodynamic model. The nonlocal model for electron interaction leads to spectral changes that heavily depend on the gap width. The field enhancement and transmitted power profiles are qualitatively identical to those computed with the LRA, whereas quantitative discrepancies arise as we explore gaps below tenths of nanometers. For these scales, we observe shifts towards the blue end of the spectrum, that is an increase in the resonant frequency ν^*, along with a decay in the maximum field enhancement π^* and increment in maximum transmission ς^*. The spreading of the electron density at the metal interface explained before is responsible for these effects, since it effectively enlarges the aperture seen by the incident EM wave. Larger effective gaps lead to an increase in transmission, as incident light now propagates through a broader aperture, and a decrease in enhancement, since there is less light confinement. In order to illustrate this behavior, the relative blueshift $\delta\nu^* = (\nu^*_{HM} - \nu^*_{LRA})/\nu^*_{LRA}$ and the ratios of maximum field enhancement $\Pi^* = \pi^*_{HM}/\pi^*_{LRA}$ and maximum transmission $\Sigma^* = \varsigma^*_{HM}/\varsigma^*_{LRA}$, are presented in Figs. 2-5e and 2-5f, where the impact of gap width on nonlocality may be readily identified. Certainly, smaller gaps exhibit large shifts, even beyond 10% for subnanometric widths, whereas the spectral response for gaps above 10 nm remains unchanged.

These effects have been observed for nanoparticles and plasmonic dimers [75,212,213,279], but have never been reported for neither annular structures nor at low THz frequencies. These results motivate the need to account for the hydrodynamic pressure in the simulation of realistic 3d plasmonic structures, since the nonlocal effects do have a substantial impact on the performance of the device for shrinking nanogaps.

2.5.6 Quantum models

Despite providing a more accurate characterization of light-metal interaction than LRA, neither the HM nor the GNOR account for quantum effects, such as quantum tunnelling. This quantum phenomena have been experimentally observed at the subnanometer scale, and have a severe impact in the properties of the plasmonic structures [45, 229, 233]. A review of the state-of-the-art quantum models for plasmonics may be found in [79, 279].

In order to replicate all the non-classical effects that can be observed experimentally, one needs to resort to *ab initio* techniques, that is modeling all the electrons in the system. Time-dependent density-functional theory (TDDFT) has been used to compute the fully quantum response of nanoparticle systems [191, 281], showing an excellent agreement with experimental observations. However, performing simulations for all the electrons in the system is computationally demanding, and can quickly become infeasible if we are interested in modeling structures above tenths of nanometers.

Recently, semi-classical approaches have been proposed to account for quantum effects within the classical electromagnetic wave equations. The class of quantum-corrected models (QCM) [69] aims to incorporate electron spill-out that has been experimentally observed for subnanometer apertures. The phenomenon of electrons escaping the metal, which is neglected by the HM, generates a current density between two closely separated metallic nanoparticles. This flow is known as electron tunnelling, and is responsible for the suppression of the field enhancement that occurs within gaps below 5 Å. Electron tunnelling is modeled by prescribing a special permittivity for the gap region, defined using quantum calculations. The QCM has been combined with the HM to encompass both quantum and nonlocal effects [70].

Another family of approaches that seek to incorporate electron spill-out to the HM is the quantum hydrodynamic theory (QHT) [44, 251]. The underlying idea is twofold: firstly we drop the assumption that only the Thomas-Fermi kinetic energy is relevant, and account for the summation of all the interacting energies in the last term of (2.8); and secondly we augment the hydrodynamic equations to simultaneously solve for the electron density profile $n_1(\overline{x}, \bar{t})$. QHT has been shown to reproduce the experimental and TDDFT results for plasmonic dimers, and is a promising method to be considered as future developments.

Chapter 3

Hybridizable discontinuous Galerkin method for plasmonics

The ability to accurately simulate surface plasmon resonances presents unique challenges. Surface plasmons involve complex geometries and a mismatch in critical length scales, which can be of several orders of magnitude. It is often necessary to use anisotropic, unstructured and curved meshes to represent the geometries, ensuring the small scales are resolved. Moreover, to minimize dispersion errors, we must seek for high order methods. As a result, the simulation of 3d EM wave propagation can be computationally demanding. To address the above issues, we use a high order unstructured finite element method known as hybridizable discontinuous Galerkin (HDG) method [47, 48, 50, 52].

We first review the HDG method for the time-harmonic Maxwell's equations proposed in [179], focusing on light-metal interaction for the local response approximation. We then formulate the HDG method for the hydrodynamic model introduced in Section 2.5.2, enabling more physically accurate simulations of plasmon resonances at the nanoscale. We provide numerical results to illustrate the features of both the local response approximation and the hydrodynamic model, as well as their performance for realistic 2d and 3d applications.

3.1 Notation

We first introduce the basic notation, operators and approximation spaces needed for the HDG method for Maxwell's equations in 3d, following [179]. We denote by $\mathcal{T}_h$ a triangulation of disjoint regular elements T that partition an open 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 $F = \partial T^+ \cap T^-$ 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 $L^2(\mathcal{D})$ being 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 $H^1(\mathcal{D})$ denote the Hilbert space with $H^1(\mathcal{D}) = \{v \in L^2(\mathcal{D}) : \int_\mathcal{D} |\nabla v|^2 < \infty\}$. We now introduce the curl-conforming space

$$\boldsymbol{H}^{\mathrm{curl}}(\mathcal{D}) = \{\mathbf{u} \in \boldsymbol{L}^2(\mathcal{D}) : \nabla \times \mathbf{u} \in \boldsymbol{L}^2(\mathcal{D})\}$$

with associated norm $\|\mathbf{u}\|^2_{\boldsymbol{H}^{\mathrm{curl}}(\mathcal{D})} = \int_{\mathcal{D}} |\mathbf{u}|^2 + |\nabla \times \mathbf{u}|^2$, as well as the div-conforming space

$$\boldsymbol{H}^{\mathrm{div}}(\mathcal{D}) = \{\mathbf{u} \in \boldsymbol{L}^2(\mathcal{D}) : \nabla \cdot \mathbf{u} \in L^2(\mathcal{D})\}$$

with associated norm $\|\mathbf{u}\|^2_{\boldsymbol{H}^{\mathrm{div}}(\mathcal{D})} = \int_{\mathcal{D}} |\mathbf{u}|^2 + |\nabla \cdot \mathbf{u}|^2$.

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

$$\begin{aligned}
W_h &= \{w \in L^2(\mathcal{D}) : w|_T \in \mathcal{P}^p(T),\ \forall T \in \mathcal{T}_h\},\\
\boldsymbol{W}_h &= \{\boldsymbol{\xi} \in \boldsymbol{L}^2(\mathcal{D}) : \boldsymbol{\xi}|_T \in [\mathcal{P}^p(T)]^3,\ \forall T \in \mathcal{T}_h\},\\
M_h &= \{\mu \in L^2(\mathcal{E}_h)\ : \mu|_F \in \mathcal{P}^p(F),\ \forall F \in \mathcal{E}_h\},\\
\boldsymbol{M}_h &= \{\boldsymbol{\mu} \in \boldsymbol{L}^2(\mathcal{E}_h)\ : \boldsymbol{\mu}|_F \in \mathcal{P}^p(F)\mathbf{t}_1 \oplus \mathcal{P}^p(F)\mathbf{t}_2,\ \forall F \in \mathcal{E}_h\},
\end{aligned}$$

where $\mathbf{t}_1$, $\mathbf{t}_2$ are vectors tangent to the face, thus naturally including the $\boldsymbol{H}^{\mathrm{curl}}$ 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 are defined in terms of $\mathbf{n} = (n_1, n_2, n_3)$, where $\mathbf{t}_1 = (-n_2/n_1, 1, 0)$ and $\mathbf{t}_2 = (-n_3/n_1, 0, 1)$ if n_1 is the largest component, and analogously for the remaining cases. Boundary conditions are included by setting $\boldsymbol{M}_h(\mathbf{u}_b) = \{\boldsymbol{\mu} \in \boldsymbol{M}_h : \mathbf{n} \times \boldsymbol{\mu} = \Pi\mathbf{u}_b \text{ on } \partial\mathcal{D}\}$ and $M_h(u_b) = \{\mu \in M_h : \mu = \Pi u_b \text{ on } \partial\mathcal{D}\}$, where $\Pi\mathbf{u}_b$ (respectively, Πu_b) is the projection of $\mathbf{u}_b$ onto $\boldsymbol{M}_h$ (respectively, u_b onto M_h).

Finally, we define the various Hermitian products for the above finite element spaces. For two arbitrary scalar functions η, ζ, its scalar product $(\eta, \zeta)_{\mathcal{D}}$ is given by the integral of $\eta\zeta^*$ on $\mathcal{D}$. Thus, the volume inner products are defined as

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

Similarly, the surface inner products are given by

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

3.2 HDG method for Maxwell's equations

We now briefly review the HDG method for time-harmonic Maxwell's equations introduced in [179], for a computational domain Ω that can be either a dielectric, a metal modeled with Drude's permittivity or a combination of both. The material variability can be encoded using a spatial dependent permittivity $\varepsilon(\mathbf{x})$.

Introducing the additional variable $\mathbf{V} = i\omega\mathbf{H}$, the original equations (2.4) are recast as: find $(\mathbf{E}, \mathbf{V}) \in \boldsymbol{H}^{\text{curl}}(\Omega) \times \boldsymbol{H}^{\text{curl}}(\Omega)$ such that

$$\mathcal{L} \begin{cases} \nabla \times \mathbf{E} - \mathbf{V} = 0, \\ \nabla \times \mathbf{V} - \omega^2 \varepsilon \mathbf{E} = 0, \end{cases} \tag{3.1}$$

along with boundary conditions

$$\mathcal{B} \begin{cases} \mathbf{n} \times \mathbf{E} \times \mathbf{n} = \mathbf{E}_D, \quad \text{on } \partial\Omega_D, \\ \mathbf{n} \times \mathbf{V} = \mathbf{n} \times \mathbf{V}_N, \quad \text{on } \partial\Omega_N, \\ \mathbf{n} \times \mathbf{V} + i\omega\sqrt{\varepsilon}\mathbf{n} \times \mathbf{E} \times \mathbf{n} = \mathbf{f}^{inc}, \quad \text{on } \partial\Omega_{SM}, \end{cases} \tag{3.2}$$

where $\mathbf{f}^{inc} = \mathbf{n} \times (\nabla \times \mathbf{E}^{inc}) + i\omega\sqrt{\varepsilon}\mathbf{n} \times \mathbf{E}^{inc} \times \mathbf{n}$. The last equation is the Silver-Müller condition, prescribed on the boundary computational domain.

We seek an approximation $(\mathbf{E}_h, \mathbf{V}_h, \widehat{\mathbf{E}}_h) \in \boldsymbol{W}_h \times \boldsymbol{W}_h \times \boldsymbol{M}_h(\mathbf{E}_D)$ such that

$$\begin{aligned} (\mathbf{V}_h, \boldsymbol{\kappa})_{\mathcal{T}_h} - (\mathbf{E}_h, \nabla \times \boldsymbol{\kappa})_{\mathcal{T}_h} - \langle \widehat{\mathbf{E}}_h, \boldsymbol{\kappa} \times \mathbf{n} \rangle_{\partial\mathcal{T}_h} &= 0, \\ (\mathbf{V}_h, \nabla \times \boldsymbol{\xi})_{\mathcal{T}_h} + \langle \widehat{\mathbf{V}}_h, \boldsymbol{\xi} \times \mathbf{n} \rangle_{\partial\mathcal{T}_h} - \omega^2(\varepsilon\mathbf{E}_h, \boldsymbol{\xi})_{\mathcal{T}_h} &= 0, \\ -\langle \mathbf{n} \times \widehat{\mathbf{V}}_h, \boldsymbol{\mu} \rangle_{\partial\mathcal{T}_h \setminus \partial\Omega_D} + \langle \widehat{\mathbf{E}}_h, \boldsymbol{\mu} \rangle_{\partial\Omega_D} - i\omega\langle \sqrt{\varepsilon}\widehat{\mathbf{E}}_h, \boldsymbol{\mu} \rangle_{\partial\Omega_{SM}} &= \langle \mathbf{F}, \boldsymbol{\mu} \rangle_{\partial\Omega}, \end{aligned} \tag{3.3}$$

holds for all $(\boldsymbol{\kappa}, \boldsymbol{\xi}, \boldsymbol{\mu}) \in \boldsymbol{W}_h \times \boldsymbol{W}_h \times \boldsymbol{M}_h(\mathbf{0})$, where $\widehat{\mathbf{E}}_h$ approximates the tangential electric field. The boundary flux is given by

$$\mathbf{F} = -\mathbf{n} \times \mathbf{V}_N \restriction_{\partial\Omega_N} - \mathbf{f}^{inc} \restriction_{\partial\Omega_{SM}}. \tag{3.4}$$

We then introduce the flux of the magnetic field as

$$\widehat{\mathbf{V}}_h = \mathbf{V}_h + \tau_t(\mathbf{E}_h - \widehat{\mathbf{E}}_h) \times \mathbf{n}, \tag{3.5}$$

where the parameter τ_t is the tangential stabilization parameter, defined globally to ensure the accuracy and stability of the HDG discretization. Based on a dimensional analysis it can be chosen as $\tau_t = \sqrt{\varepsilon}\omega$. Introducing (3.5) in (3.3) and integrating by parts, we arrive to the final HDG discretization for Maxwell's equations, namely

$$(\mathbf{V}_h, \boldsymbol{\kappa})_{\mathcal{T}_h} - (\mathbf{E}_h, \nabla \times \boldsymbol{\kappa})_{\mathcal{T}_h} - \langle \widehat{\mathbf{E}}_h, \boldsymbol{\kappa} \times \mathbf{n} \rangle_{\partial\mathcal{T}_h} = 0, \tag{3.6a}$$

$$(\nabla \times \mathbf{V}_h, \boldsymbol{\xi})_{\mathcal{T}_h} + \langle \tau_t(\mathbf{E}_h - \widehat{\mathbf{E}}_h), \mathbf{n} \times \boldsymbol{\xi} \times \mathbf{n} \rangle_{\partial\mathcal{T}_h} - \omega^2(\varepsilon\mathbf{E}_h, \boldsymbol{\xi})_{\mathcal{T}_h} = 0, \tag{3.6b}$$

$$-\langle \mathbf{n} \times \mathbf{V}_h + \tau_t \mathbf{E}_h, \boldsymbol{\mu} \rangle_{\partial\mathcal{T}_h \backslash \partial\Omega_D} + \langle \widetilde{\tau}_t \widehat{\mathbf{E}}_h, \boldsymbol{\mu} \rangle_{\partial\mathcal{T}_h} = \langle \mathbf{F}, \boldsymbol{\mu} \rangle_{\partial\Omega}. \tag{3.6c}$$

where

$$\widetilde{\tau}_t = \begin{cases} \tau_t, & \text{on } \partial\mathcal{T}_h \backslash \partial\Omega_D \cup \partial\Omega_{SM} \\ \tau_t - i\omega\sqrt{\varepsilon}, & \text{on } \partial\Omega_{SM} \\ 1, & \text{on } \partial\Omega_D \end{cases} \tag{3.7}$$

General implementation details may be found in [179,224]. However, in Section 4.2.2 we review the implementation and solution guidelines for a more generic formulation of Maxwell's equations.

3.2.1 Numerical results

In this section, we present two plasmonic applications to demonstrate the performance and effectiveness of the HDG method.

Nanoslit

The problem under consideration is an infinitely long 2d aperture in a thin gold film that, when illuminated from below with a plane wave polarized along the perpendicular direction of the slit, is able to transfer light from the lower part of the film to the upper part of the film. This extraordinary optical transmission is produced by the combination of surface plasmons

on both surfaces of the slit and cavity resonances. This structure has been extensively studied both theoretically and experimentally, and we refer the reader to the review paper [81] and the references therein for a detailed discussion.

A sketch of the nanoslit that will be simulated is shown in Fig. 3-1a, and the gapsizes considered are $w = 2$, 5, 10 and 20 nm. The thickness of the gold film is taken equal to 150 nm. Furthermore, we focus on the FIR regime, for which field enhancements of several orders of magnitude have been reported in literature [39, 235].

The substrate for the gold film is silica SiO_2, which is transparent at FIR, with a constant refractive index of $n_{SiO_2} = 1.96$ given by Naftaly *et al.* [166]. For this low frequencies, the optical constants of gold are adopted from Ordal *et al.* [192, 193], and we consider the nominal values $\hbar\omega_p = 9.02$ eV, $\hbar\gamma = 0.02678$ eV and $\varepsilon_\infty = 1$. The dielectric material used for the gap is alumina Al_2O_3, and we use the thickness-dependent values for low frequencies experimentally observed in [92] by Groner *et al.*, and reproduced in Fig. 3-1b (left). For the gaps under study, these values correspond to $\varepsilon_{Al_2O_3} = \{3, 4.4, 5.5, 6.4\}$. This model for the permittivity of Al_2O_3 has been used for FIR simulations [40, 198] and shown to produce good qualitative agreeement with experimental results.

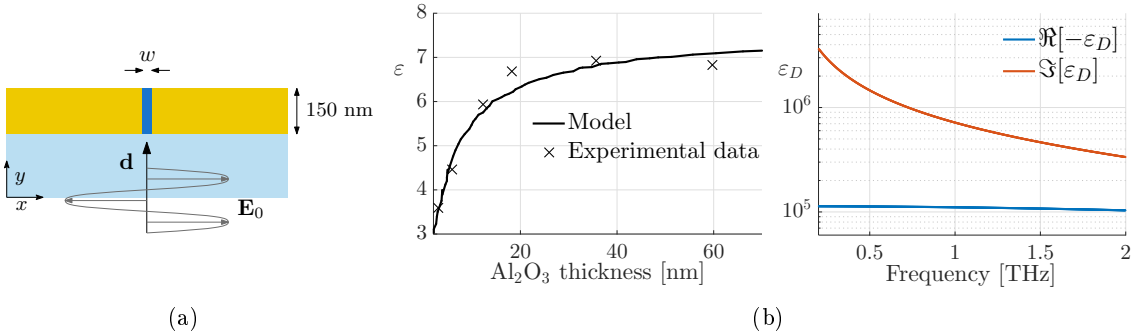


Figure 3-1: (a) Schematic of gold nanoslit (b) Left: thickness-dependent permittivity for Al_2O_3 adapted from [92]. Right: Real and imaginary part of Drude permittivity for gold.

The numerical simulation of the nanoslit in the FIR, despite being in 2d, can be quite challenging due to the tremendous disparity in length scales present in the problem: the thickness of the film is in the order of tenths of nanometers, the gap widths are in the nanometer scale and the relevant wavelengths span from hundreds of microns to millimeters. Moreover, the simulation technique needs to accommodate discontinuities in the material properties of the different components, which can again differ by several orders of magnitude.

For instance, the Drude model for gold at low THz frequencies is given by the permittivity coefficient shown in Fig. 3-1b (right), hence almost behaving as a perfect electric conductor. The Drude permittivity is several orders of magnitude larger than the constants for silica and alumina, thus posing additional simulation challenges.

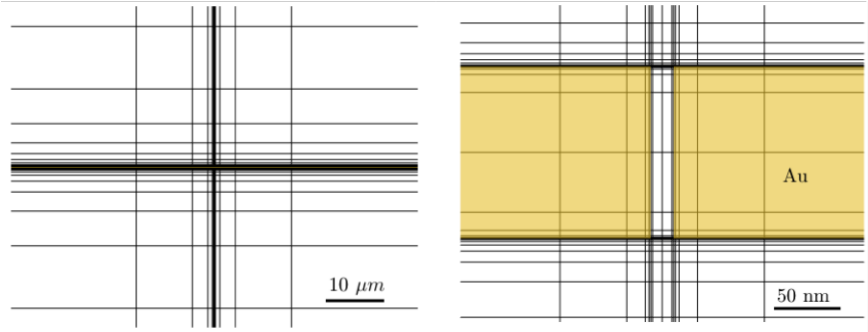
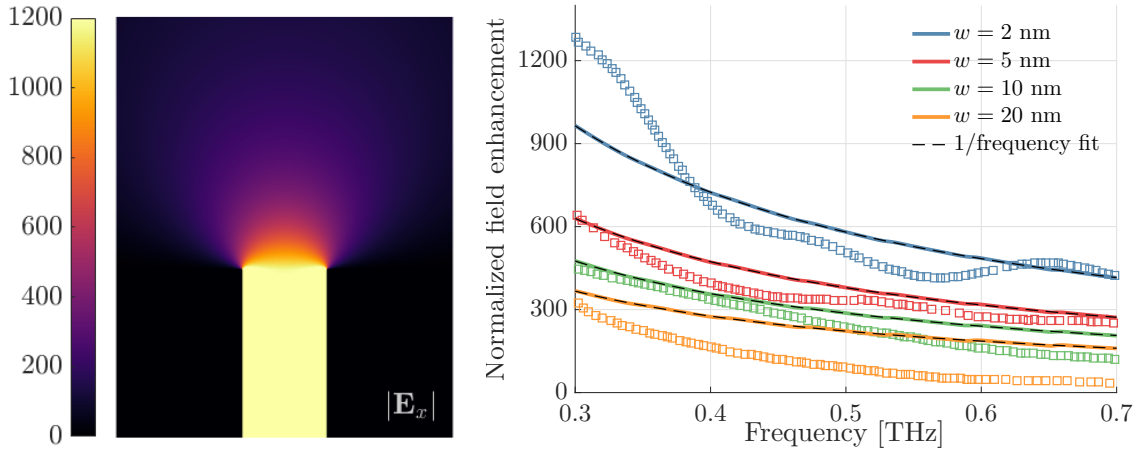


Figure 3-2: Nanoslit discretization with two different zooms, with gold film highlighted. Discretization used in calculations has 4 times more elements.

For this simulation, we set the computational domain to be a square of 2.5 mm$\times$ 0.5 mm, and prescribe Silver-Müller conditions on the top/bottom boundaries and periodicity on the lateral boundaries. The size of the computational domain is chosen such that the location of the radiating boundaries is far enough so that it has no significant effect on the solution. The domain is discretized with an anisotropic mesh of 20K cubic quadrangular elements, ensuring that greater resolution is achieved near the aperture, see Fig. 3-2, with element sizes ranging from 10 mm to 0.1 nm. The numerical accuracy is verified by carrying out grid convergence studies on consecutively refined meshes until the relative error for the field enhancement of the 2 nm gap is below 0.05%.

In Fig. 3-3a, we show $|\mathbf{E}_x|$ at the upper surface of the slit to better appreciate the extraordinary field enhancement that occurs within the aperture, for an incident frequency of 0.3 THz and gap width of 2 nm. This excitation is constant along the slit, and it decays exponentially fast just nanometers away from the gold film. Moreover, the metal acts almost as a perfect mirror, since only allows minimal penetration of the impinging EM wave. For the remaining gaps the solution field exhibits the same pattern but downscaled, since in all cases the field enhancement is constant along the aperture.

Finally, in Fig. 3-3b we perform a frequency sweep study for the various gaps and report the field enhancement attained, which varies inversely with the frequency. The field enhance-



(a) (b)

Figure 3-3: (a) Field enhancement for 2 nm wide slit illuminated at 0.3 THz. (b) Normalized field enhancement for different gaps and fitted 1/frequency curves. Experimental data is shown with square dots.

ment is normalized by the electric field amplitude for a silica-air interface, without the metal or gap. This normalization ensures the value of the field enhancement does not depend on the propagation direction, *i.e.* air to silica or silica to air. Results from HDG simulations are compared to the experimental data in [39], reproduced in Fig. 3-3b with square dots. There is a good qualitative agreement between data and simulations, specially for 5 and 10 nm slits, although the field enhancement is often overestimated with respect to the measured values. This mismatch could be caused by a poor characterization of the permittivity constant for Al_2O_3, since the alumina film on the fabricated slit may be subject to thermal annealing or exhibit voids, which is reported to impact its effective refractive index [266]. Alternatively, numerical simulation introduces artifacts such as sharp corners that lead to larger enhancements, whereas in fabrication corners are rounded, thus increasing the gap size seen by the incident wave.

Periodic annular structure

This next example consists of a 3d periodic annular structure shown in Fig. 3-4. We are interested in the low THz regime (0.2 to 2 THz), which requires diameters on the order of microns to excite resonances. The thickness of the gold film is taken equal to 150 nm over a SiO_2 substrate, and the annular gaps are filled with Al_2O_3. The optical constants of the materials are the same as in the nanoslit example above. For this problem, we choose the

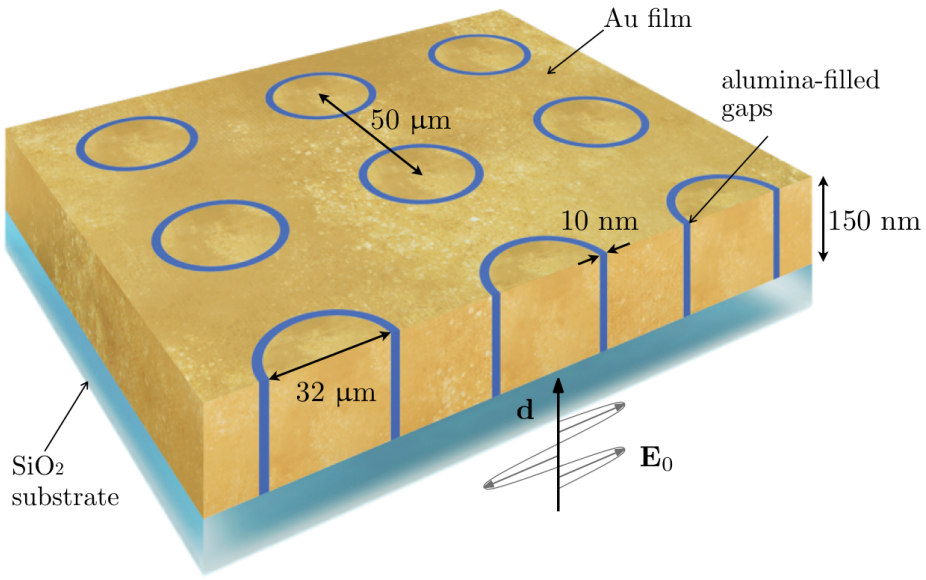


Figure 3-4: Schematic diagram of thin gold film on silica substrate patterned with periodic square array of alumina gaps under plane wave THz illumination.

diameter of the rings to be 32 microns with square periodicity of 50 μm, and a gap width of 10 nm. Note the problem may be further reduced by capitalizing the symmetries of the lattice, hence we only need to solve for one quadrant of the ring structure as indicated in Fig. 3-5a.

From the computational perspective, simulating this structure is quite challenging. Firstly, the required meshes need not only be curved to represent the annular geometry, but also must exhibit high anisotropy to properly capture the tight field confinement that occurs in the vicinity of the gap. In addition, we must accommodate the different length scales of the problem, since for the lowest frequencies the wavelength-gap ratio can be of 5 orders of magnitude. The HDG method has already been proved to successfully tackle this particular problem [198]. Nonetheless, we review it here as a building block for more complicated examples that will be analyzed in this dissertation.

The high order discretization consists of 1.6K hexahedral quartic elements, and is devised as a 2.5d mesh by extruding in the z-direction the 2d curved mesh shown in Fig. 3-5b. Similarly as in the nanoslit case, the hexagonal elements in the vertical direction are smaller close to the upper and lower surfaces of the gold film, and gradually increase as we separate from the metal. The radiation conditions are prescribed at 30 microns for the glass substrate and 30 microns for air, ensuring there is no numerical interaction between the boundary and

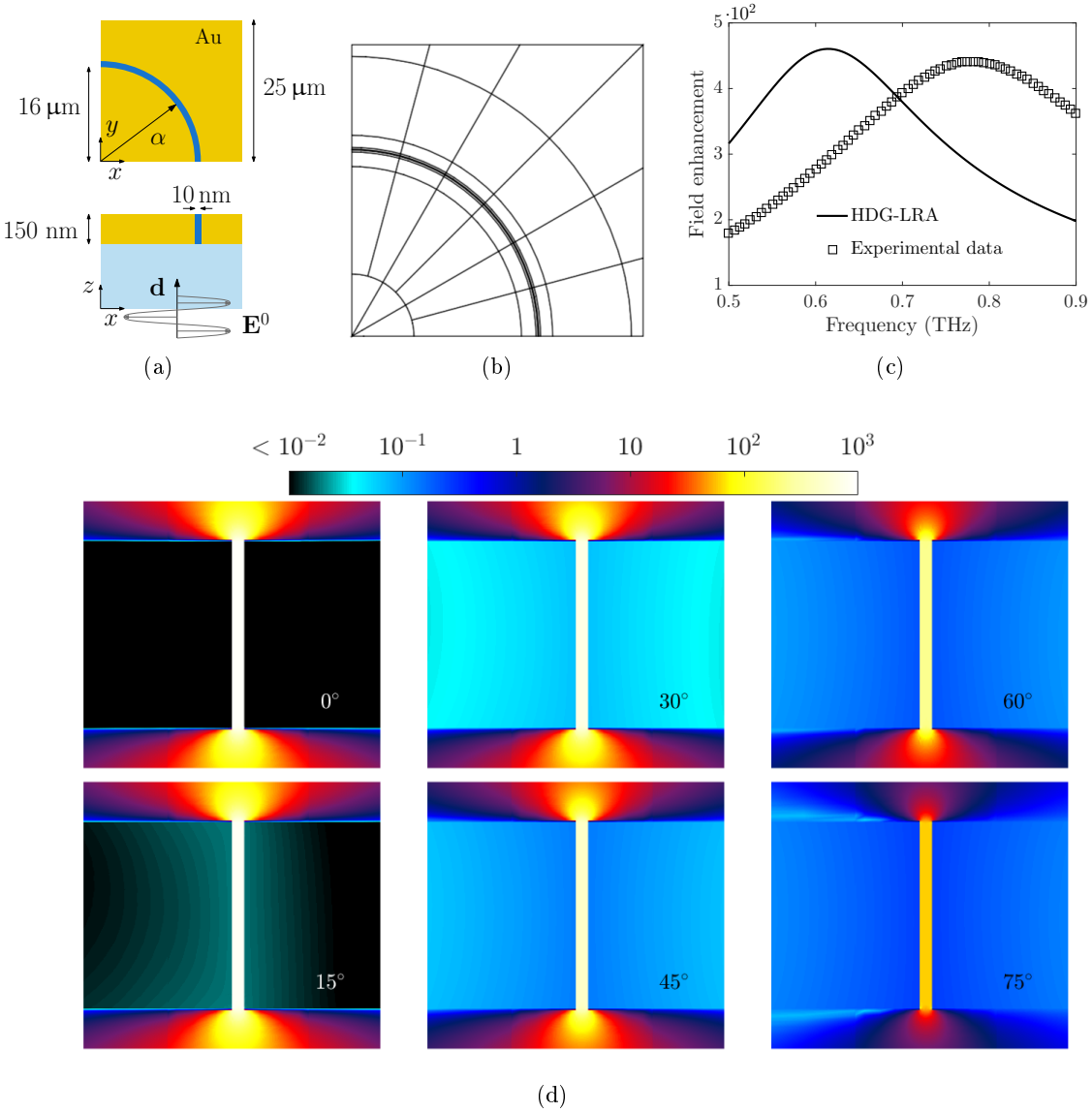


Figure 3-5: (a) Schematic of unit computational cell for periodic annular array. (b) 2d curved mesh. (c) Field enhancement curve for LRA compared with experimental data. (d) Slices in the angular direction of solution field $|\mathbf{E}_x|$ at the aperture for 0.6 THz, shown in logarithmic scale.

the extraordinary optical transmission that occurs in the ring. Under these conditions, the reduced problem for the traces of the tangential component of the electric field consists of 270K unknowns, and we use MATLAB's built-in sparse direct solver to compute the solution.

The magnitude of the x-component of the electric field for an x-polarized incident electromagnetic wave is shown in Fig. 3-5d, for the planes $y/x = \tan\alpha$ specified in Fig. 3-5a. The pattern of extreme confinement is similar to that of the nanoslit, that is, the enhancement is constant throughout the annular aperture for all angles, reaching a maximum at $\alpha = 0°$

and decreasing as the angle approaches 90°. Again, the penetration in the metal is minimal, although it is interesting to observe that the further we shift from the resonance in the angular direction, the more the electric field propagates inside the metal. Lastly, we perform a frequency sweep to detect the peak resonance and frequency, which occurs at 0.6 THz. These results can be compared with experimental data [198], whereby we find that the simulations qualitatively agree with the experiments, but there is a significant blueshift in the resonant frequency. In the remainder of the dissertation we will attempt to explain some possible factors for this mismatch, beginning by introducing nonlocality in our simulation models.

3.3 HDG method for the hydrodynamic model

We now develop an HDG method to numerically solve the Maxwell equations with the hydrodynamic model (2.11) for a metallic computational domain $\overline{\Omega}$, which will serve as a building block towards more complicated scenarios. We introduce additional variables $\mathbf{V} = i\omega\mathbf{H}$, $U = \nabla \cdot \mathbf{J}$ and rewrite system (2.11) as a first order system of equations in $\overline{\Omega}$:

$$\overline{\mathcal{L}} \begin{cases} \nabla \times \mathbf{E} - \mathbf{V} = 0, \\ \beta^2 \nabla U + \omega(\omega + i\gamma)\mathbf{J} - i\omega\omega_p^2 \mathbf{E} = 0, \\ \nabla \times \mathbf{V} - \omega^2 \varepsilon_\infty \mathbf{E} - i\omega \mathbf{J} = 0, \\ U - \nabla \cdot \mathbf{J} = 0. \end{cases} \tag{3.8}$$

The additional variable U is related to the induced free charge density in the metal as $\rho_f = -i\omega U$. The system above is completed with boundary conditions

$$\begin{aligned} &\mathbf{n} \times \mathbf{E} \times \mathbf{n} = \mathbf{E}_D, \quad \text{on } \partial\overline{\Omega}_D, \\ &\mathbf{n} \times \mathbf{V} = \mathbf{n} \times \mathbf{V}_N, \quad \text{on } \partial\overline{\Omega}_N, \\ &\mathbf{n} \cdot \mathbf{J} = 0, \quad \text{on } \partial\overline{\Omega}. \end{aligned}$$

The system of equations in (3.8) has a solution $(\mathbf{E}, \mathbf{V}, \mathbf{J}, U) \in \mathbf{H}^{\text{curl}}(\overline{\Omega}) \times \mathbf{H}^{\text{curl}}(\overline{\Omega}) \times \mathbf{H}^{\text{div}}(\overline{\Omega}) \times H^1(\overline{\Omega})$, and we seek an approximation

$(\mathbf{V}_h, \mathbf{E}_h, \mathbf{J}_h, U_h, \widehat{\mathbf{E}}_h, \widehat{U}_h) \in \boldsymbol{W}_h \times \boldsymbol{W}_h \times \boldsymbol{W}_h \times W_h \times \boldsymbol{M}_h(\mathbf{E}_D) \times M_h$ such that

$$\begin{aligned}
(\mathbf{V}_h, \boldsymbol{\kappa})_{\mathcal{T}_h} - (\mathbf{E}_h, \nabla \times \boldsymbol{\kappa})_{\mathcal{T}_h} - \langle \widehat{\mathbf{E}}_h, \boldsymbol{\kappa} \times \mathbf{n} \rangle_{\partial \mathcal{T}_h} &= 0, \\
-\beta^2 (U_h, \nabla \cdot \boldsymbol{\eta})_{\mathcal{T}_h} + \beta^2 \langle \widehat{U}_h, \boldsymbol{\eta} \cdot \mathbf{n} \rangle_{\partial \mathcal{T}_h} + \omega(\omega + i\gamma)(\mathbf{J}_h, \boldsymbol{\eta})_{\mathcal{T}_h} - i\omega\omega_p^2 (\mathbf{E}_h, \boldsymbol{\eta})_{\mathcal{T}_h} &= 0, \\
(\mathbf{V}_h, \nabla \times \boldsymbol{\xi})_{\mathcal{T}_h} + \langle \widehat{\mathbf{V}}_h, \boldsymbol{\xi} \times \mathbf{n} \rangle_{\partial \mathcal{T}_h} - \omega^2 (\varepsilon_\infty \mathbf{E}_h, \boldsymbol{\xi})_{\mathcal{T}_h} - i\omega (\mathbf{J}_h, \boldsymbol{\xi})_{\mathcal{T}_h} &= 0, \\
(U_h, \zeta)_{\mathcal{T}_h} - \langle \widehat{\mathbf{J}}_h \cdot \mathbf{n}, \zeta \rangle_{\partial \mathcal{T}_h} + (\mathbf{J}_h, \nabla \zeta)_{\mathcal{T}_h} &= 0, \\
-\langle \mathbf{n} \times \widehat{\mathbf{V}}_h, \boldsymbol{\mu} \rangle_{\partial \mathcal{T}_h \setminus \partial \overline{\Omega}_D} + \langle \widehat{\mathbf{E}}_h, \boldsymbol{\mu} \rangle_{\partial \overline{\Omega}_D} - \langle \mathbf{F}, \boldsymbol{\mu} \rangle_{\partial \overline{\Omega}} &= 0, \\
\langle \widehat{\mathbf{J}}_h \cdot \mathbf{n}, \theta \rangle_{\partial \mathcal{T}_h} &= 0,
\end{aligned} \tag{3.9}$$

holds for all $(\boldsymbol{\kappa}, \boldsymbol{\eta}, \boldsymbol{\xi}, \zeta, \boldsymbol{\mu}, \theta) \in \boldsymbol{W}_h \times \boldsymbol{W}_h \times \boldsymbol{W}_h \times W_h \times \boldsymbol{M}_h(\mathbf{0}) \times M_h$, where $\widehat{\mathbf{E}}_h$ approximates the tangential field of $\mathbf{E}_h$, and $\widehat{U}_h$ approximates the trace of U_h. We then introduce the trace of the magnetic field as (3.5) and the trace of the electric current field as

$$\widehat{\mathbf{J}}_h = \mathbf{J}_h - \tau_n (U_h - \widehat{U}_h)\mathbf{n}. \tag{3.10}$$

The parameters τ_t, τ_n are the stabilization parameters, defined globally to ensure the accuracy and stability of the HDG discretization. We propose to select τ_n as the inverse of δ, that is $\tau_n = \omega_p/\beta$. Intuitively, we enforce a stabilization inversely proportional to the penetration distance of the EM wave into the metal. This choice leads to numerically stable solutions even in the presence of tightly localized fields in the metal-dielectric interface.

For compactness, we borrow the definitions of boundary flux $\mathbf{F}$ and the generic stabilization constant $\widetilde{\tau}_t$ given in (3.4) and (3.7), respectively. Introducing (3.5) and (3.10) in (3.9) and integrating by parts, we write the final HDG discretization of the hydrodynamic model for metals as

$$
\begin{aligned}
(\mathbf{V}_h, \boldsymbol{\kappa})_{\mathcal{T}_h} - (\mathbf{E}_h, \nabla \times \boldsymbol{\kappa})_{\mathcal{T}_h} - \langle \widehat{\mathbf{E}}_h, \boldsymbol{\kappa} \times \mathbf{n} \rangle_{\partial \mathcal{T}_h} &= 0, \\
-\beta^2 (U_h, \nabla \cdot \boldsymbol{\eta})_{\mathcal{T}_h} + \beta^2 \langle \widehat{U}_h, \boldsymbol{\eta} \cdot \mathbf{n} \rangle_{\partial \mathcal{T}_h} + \omega(\omega + i\gamma)(\mathbf{J}_h, \boldsymbol{\eta})_{\mathcal{T}_h} - i\omega\omega_p^2 (\mathbf{E}_h, \boldsymbol{\eta})_{\mathcal{T}_h} &= 0, \\
(\nabla \times \mathbf{V}_h, \boldsymbol{\xi})_{\mathcal{T}_h} + \langle \tau_t [\mathbf{E}_h - \widehat{\mathbf{E}}_h], \mathbf{n} \times \boldsymbol{\xi} \times \mathbf{n} \rangle_{\partial \mathcal{T}_h} - \omega^2 (\varepsilon_\infty \mathbf{E}_h, \boldsymbol{\xi})_{\mathcal{T}_h} - i\omega (\mathbf{J}_h, \boldsymbol{\xi})_{\mathcal{T}_h} &= 0, \\
-(\nabla \cdot \mathbf{J}_h, \zeta)_{\mathcal{T}_h} + (U_h, \zeta)_{\mathcal{T}_h} + \langle \tau_n U_h, \zeta \rangle_{\partial \mathcal{T}_h} - \langle \tau_n \widehat{U}_h, \zeta \rangle_{\partial \mathcal{T}_h} &= 0, \\
-\langle \mathbf{n} \times \mathbf{V}_h + \tau_t \mathbf{E}_h, \boldsymbol{\mu} \rangle_{\partial \mathcal{T}_h \setminus \partial \overline{\Omega}_D} + \langle \widetilde{\tau}_t \widehat{\mathbf{E}}_h, \boldsymbol{\mu} \rangle_{\partial \mathcal{T}_h} - \langle \mathbf{F}, \boldsymbol{\mu} \rangle_{\partial \overline{\Omega}} &= 0, \\
\langle \mathbf{J}_h \cdot \mathbf{n}, \theta \rangle_{\partial \mathcal{T}_h} - \langle \tau_n U_h, \theta \rangle_{\partial \mathcal{T}_h} + \langle \tau_n \widehat{U}_h, \theta \rangle_{\partial \mathcal{T}_h} &= 0.
\end{aligned}
\tag{3.11}
$$

The first four equations represent the weak formulation of equations (3.8), whereas the last two equations enforce zero jump in the tangential component of $\mathbf{V}_h$ and in the normal component of $\mathbf{J}_h$ respectively, along with the appropriate boundary conditions. Proofs regarding conservation, consistency and well-posedness are given in Appendix A.

3.3.1 Implementation

We now describe the relevant implementation steps of the HDG method for the hydrodynamic model. The system of equations in (3.11) is rewritten for convenience in terms of several bilinear forms. The weak formulation reads: find $(\mathbf{E}_h, \mathbf{V}_h, \mathbf{J}_h, U_h, \widehat{\mathbf{E}}_h, \widehat{U}_h) \in \boldsymbol{W}_h \times \boldsymbol{W}_h \times \boldsymbol{W}_h \times W_h \times \boldsymbol{M}_h(\mathbf{E}_D) \times M_h$ such that

$$
\begin{aligned}
\mathcal{A}(\mathbf{V}_h, \boldsymbol{\kappa}) - \mathcal{B}(\mathbf{E}_h, \boldsymbol{\kappa}) - \mathcal{C}(\widehat{\mathbf{E}}_h, \boldsymbol{\kappa}) &= 0, \\
\omega(\omega + i\gamma)\mathcal{A}(\mathbf{J}_h, \boldsymbol{\eta}) - i\omega\omega_p^2 \mathcal{A}(\mathbf{E}_h, \boldsymbol{\eta}) - \beta^2 \mathcal{P}(U_h, \boldsymbol{\eta}) + \beta^2 \mathcal{O}(\widehat{U}_h, \boldsymbol{\eta}) &= 0, \\
\mathcal{B}(\boldsymbol{\xi}, \mathbf{V}_h) - i\omega \mathcal{A}(\mathbf{J}_h, \boldsymbol{\xi}) + \mathcal{D}(\mathbf{E}_h, \boldsymbol{\xi}) - \omega^2 \mathcal{A}_\varepsilon(\mathbf{E}_h, \boldsymbol{\xi}) - \mathcal{E}(\widehat{\mathbf{E}}_h, \boldsymbol{\xi}) &= 0, \\
-\mathcal{P}(\zeta, \mathbf{J}_h) + \mathcal{H}(U_h, \zeta) - \mathcal{N}(\widehat{U}_h, \zeta) &= 0, \\
-\mathcal{R}(\mathbf{V}_h, \boldsymbol{\mu}) - \mathcal{L}(\mathbf{E}_h, \boldsymbol{\mu}) + \mathcal{M}(\widehat{\mathbf{E}}_h, \boldsymbol{\mu}) &= \mathcal{F}(\boldsymbol{\mu}), \\
\mathcal{O}(\theta, \mathbf{J}_h) - \mathcal{N}(\theta, U_h) + \mathcal{T}(\widehat{U}_h, \theta) &= 0,
\end{aligned}
\tag{3.12}
$$

holds for all $(\boldsymbol{\kappa}, \boldsymbol{\eta}, \boldsymbol{\xi}, \zeta, \boldsymbol{\mu}, \theta) \in \boldsymbol{W}_h \times \boldsymbol{W}_h \times \boldsymbol{W}_h \times W_h \times \boldsymbol{M}_h(\mathbf{0}) \times M_h$. The bilinear forms are given by

$$
\begin{aligned}
&\mathcal{A}(\mathbf{V}, \boldsymbol{\kappa}) = (\mathbf{V}, \boldsymbol{\kappa})_{\mathcal{T}_h}, && \mathcal{A}_\varepsilon(\mathbf{E}, \boldsymbol{\xi}) = (\varepsilon_\infty \mathbf{E}, \boldsymbol{\xi})_{\mathcal{T}_h}, \\
&\mathcal{B}(\mathbf{E}, \boldsymbol{\kappa}) = (\mathbf{E}, \nabla \times \boldsymbol{\kappa})_{\mathcal{T}_h}, && \mathcal{C}(\widehat{\mathbf{E}}, \boldsymbol{\kappa}) = \langle \widehat{\mathbf{E}}, \boldsymbol{\kappa} \times \mathbf{n} \rangle_{\partial \mathcal{T}_h}, \\
&\mathcal{P}(U, \boldsymbol{\eta}) = (U, \nabla \cdot \boldsymbol{\eta})_{\mathcal{T}_h}, && \mathcal{O}(\widehat{U}, \boldsymbol{\eta}) = \langle U, \boldsymbol{\eta} \cdot \mathbf{n} \rangle_{\mathcal{T}_h}, \\
&\mathcal{D}(\mathbf{E}, \boldsymbol{\xi}) = \langle \tau_t \mathbf{E}, \mathbf{n} \times \boldsymbol{\xi} \times \mathbf{n} \rangle_{\partial \mathcal{T}_h}, && \mathcal{E}(\widehat{\mathbf{E}}, \boldsymbol{\xi}) = \langle \tau_t \widehat{\mathbf{E}}, \boldsymbol{\xi} \rangle_{\partial \mathcal{T}_h}, \\
&\mathcal{H}(U, \zeta) = (U, \zeta)_{\mathcal{T}_h} + \langle \tau_n U, \zeta \rangle_{\partial \mathcal{T}_h}, && \mathcal{N}(\widehat{U}, \zeta) = \langle \tau_n \widehat{U}, \zeta \rangle_{\partial \mathcal{T}_h}, \\
&\mathcal{R}(\mathbf{V}, \boldsymbol{\mu}) = \langle \mathbf{n} \times \mathbf{V}, \boldsymbol{\mu} \rangle_{\partial \mathcal{T}_h \setminus \partial \overline{\Omega}_D}, && \mathcal{L}(\mathbf{E}, \boldsymbol{\mu}) = \langle \tau_t \mathbf{E}, \boldsymbol{\mu} \rangle_{\partial \mathcal{T}_h \setminus \partial \overline{\Omega}_D}, \\
&\mathcal{M}(\widehat{\mathbf{E}}, \boldsymbol{\mu}) = \langle \widetilde{\tau}_t \widehat{\mathbf{E}}, \boldsymbol{\mu} \rangle_{\partial \mathcal{T}_h}, && \mathcal{T}(\widehat{U}, \theta) = \langle \tau_n \widehat{U}, \theta \rangle_{\partial \mathcal{T}_h}, \\
&\mathcal{F}(\boldsymbol{\mu}) = \langle \mathbf{F}, \boldsymbol{\mu} \rangle_{\partial \overline{\Omega}}.
\end{aligned}
$$

The discretization of the system of equations in (3.12) gives rise to the following matrix equation

$$
\begin{bmatrix}
\mathbb{A} & 0 & -\mathbb{B} & 0 & -\mathbb{C} & 0 \\
0 & \omega(\omega + i\gamma)\mathbb{A} & -i\omega\omega_p^2\mathbb{A} & -\beta^2\mathbb{P} & 0 & \beta^2\mathbb{O} \\
\mathbb{B}^T & -i\omega\mathbb{A} & \mathbb{D} - \omega^2\mathbb{A}_\varepsilon & 0 & -\mathbb{E} & 0 \\
0 & -\mathbb{P}^T & 0 & \mathbb{H} & 0 & -\mathbb{N} \\
-\mathbb{R} & 0 & -\mathbb{L} & 0 & -\mathbb{M} & 0 \\
0 & -\mathbb{O}^T & 0 & -\mathbb{N}^T & 0 & \mathbb{T}
\end{bmatrix}
\begin{bmatrix}
\underline{\mathbf{V}} \\ \underline{\mathbf{J}} \\ \underline{\mathbf{E}} \\ \underline{U} \\ \widehat{\underline{\mathbf{E}}} \\ \widehat{\underline{U}}
\end{bmatrix}
=
\begin{bmatrix}
0 \\ 0 \\ 0 \\ 0 \\ \mathbf{F} \\ 0
\end{bmatrix}
$$

where $\underline{\mathbf{E}}$, $\underline{\mathbf{V}}$, $\underline{\mathbf{J}}$, $\underline{U}$, $\widehat{\underline{\mathbf{E}}}$, $\widehat{\underline{U}}$ are vectors containing the values of the corresponding fields at the degrees of freedom defined by the discretization $\mathcal{T}_h$. The system above, however, is never formed in implementation. Instead, we invoke discontinuity of the approximation spaces to hybridize the linear system. That is, we locally eliminate the degrees of freedom of $\boldsymbol{\Upsilon} = (\underline{\mathbf{V}}, \underline{\mathbf{J}}, \underline{\mathbf{E}}, \underline{U})$, or local unknowns, and express them as a function of only the degrees of freedom of the approximate traces $\widehat{\boldsymbol{\Upsilon}} = [\widehat{\underline{\mathbf{E}}}, \widehat{\underline{U}}]$, or global unknowns. The relation between

global and local uknowns $\boldsymbol{\Upsilon} = \mathbb{Z}\widehat{\boldsymbol{\Upsilon}}$, defined at the element level, takes the form

$$\begin{bmatrix} \underline{\mathbf{V}} \\ \underline{\mathbf{J}} \\ \underline{\mathbf{E}} \\ \underline{U} \end{bmatrix} = \begin{bmatrix} \mathbb{A} & 0 & -\mathbb{B} & 0 \\ 0 & \omega(\omega + i\gamma)\mathbb{A} & -i\omega\omega_p^2\mathbb{A} & -\beta^2\mathbb{P} \\ \mathbb{B}^T & -i\omega\mathbb{A} & \mathbb{D} - \omega^2\mathbb{A}_\varepsilon & 0 \\ 0 & -\mathbb{P}^T & 0 & \mathbb{H} \end{bmatrix}^{-1} \begin{bmatrix} \mathbb{C} & 0 \\ 0 & -\beta^2\mathbb{O} \\ \mathbb{E} & 0 \\ 0 & \mathbb{N} \end{bmatrix} \widehat{\boldsymbol{\Upsilon}}. \tag{3.13}$$

In addition, the inverse can be computed efficiently since it is block diagonal, due to the choice of approximation spaces. This elimination of degrees of freedom renders a linear system that involves only the global degrees of freedom, defined at the discretization faces. Indeed, we eliminate the local unknowns – 10 components defined in the high-order volume nodes– and solve only for the global unknowns – 3 components defined in the high-order face nodes– hence drastically reducing the size of the linear system that must be solved. This is one of the most attractive features of the HDG method. Finally, the system involving only the global unknowns is given by

$$\left(\begin{bmatrix} -\mathbb{M} & 0 \\ 0 & \mathbb{T} \end{bmatrix} + \begin{bmatrix} -\mathbb{R} & 0 & -\mathbb{L} & 0 \\ 0 & -\mathbb{O}^T & 0 & -\mathbb{N}^T \end{bmatrix} \mathbb{Z} \right) \widehat{\boldsymbol{\Upsilon}} = \begin{bmatrix} \mathbf{F} \\ 0 \end{bmatrix}.$$

This procedure characterizes the solution to (3.11) in terms of $\widehat{\mathbf{E}}_h$ and $\widehat{U}_h$. If $\mathbb{Z}$ is stored element-wise the local volume variables can be recovered in parallel through (3.13) incurring a small cost.

3.3.2 Metal-dielectric coupling

In this section, we discuss the coupling of the HDG formulations in (3.6) and (3.11) for cases where we consider a metal $\overline{\Omega}$ described by the hydrodynamic model embedded in a dielectric medium Ω, as shown in Fig. 3-6 (left). This is a common situation where, for instance, the metallic nanostructure scatters an incident plane wave. The boundaries of the computational domain Ω represent the truncation of the infinite space, where radiation is imposed either with Silver-Müller conditions or PMLs, and must be placed far away from the scatterer.

We now have two subdomains where different governing equations. The solution within the

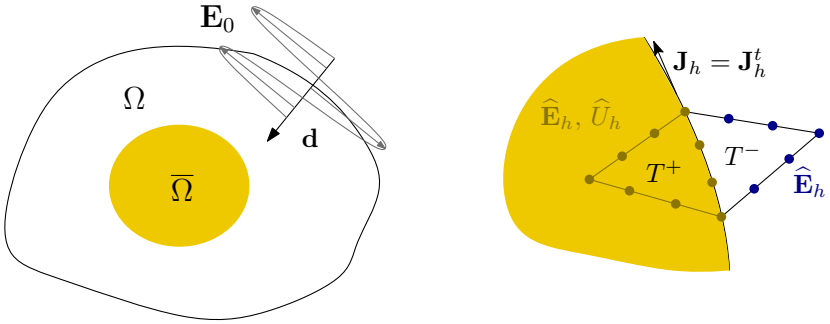


Figure 3-6: Left: Metallic structure $\overline{\Omega}$ embedded in dielectric Ω illuminated by plane wave. Right: Detail of metal-dielectric interface with global degrees of freedom.

metallic structure is governed by

$$\begin{aligned} \overline{\mathcal{L}} &= 0, \quad \text{in } \overline{\Omega}, \\ \mathbf{J} \cdot \mathbf{n} &= 0, \quad \text{on } \partial\overline{\Omega}, \end{aligned}$$

whose HDG discretization is (3.11). Note that the no electron spill-out condition enforces that the electric current at the metallic interface is tangential $\mathbf{J}_h = \mathbf{J}_h^t$. Conversely, the response in the dielectric is given by

$$\begin{aligned} \mathcal{L} &= 0, \quad \text{in } \Omega, \\ \mathcal{B} &= 0, \quad \text{on } \partial\Omega, \end{aligned}$$

with weak HDG formulation given by (3.6).

In addition, we need to impose a compatibility condition to stitch the subdomains together. For any two elements T^+, T^- that satisfy $T^+ \cap T^- \in \partial\overline{\Omega}$, see Fig. 3-6 (right), we enforce continuity of the tangential component of the trace of the magnetic field $[\![\mathbf{n} \times \widehat{\mathbf{V}}_h]\!] = 0$ at the interface. Furthermore, since the traces are single-valued across inter-element boundaries, the global degrees of freedom on the faces $F \in \partial T^-$ have two $\{\widehat{\mathbf{E}}_h\}$ and three $\{\widehat{\mathbf{E}}_h, \widehat{U}_h\}$ components for $F \notin \partial\overline{\Omega}$ and $F \in \partial\overline{\Omega}$, respectively. Thus, the assembly of the global matrix needs to account for the global compatibility condition and the different number of global components.

3.3.3 Convergence test

In this section, we perform numerical tests to examine the convergence and accuracy of the HDG method for the HM introduced above. To that end, we solve (3.8) in a square domain

$\overline{\Omega} = (0,\pi)^2$ with $\varepsilon_\infty = 2$. In addition, we set $\omega = \omega_p = 1$, $\gamma = 0$ and $\beta^2 = 0.5$ and select boundary data $\mathbf{E}_D$ such that the problem has the following exact solution

$$\mathbf{E} = (\cos x - i \sin y, \cos y - i \sin x), \quad \mathbf{J} = (\sin y + 2i \cos x, \sin x + 2i \cos y).$$

We define the error in the L^2, $\mathbf{H}^{\mathrm{div}}(\mathcal{T}_h)$ and $\mathbf{H}^{\mathrm{curl}}(\mathcal{T}_h)$ norms as

Degree	**Mesh**	$\|\mathbf{E}-\mathbf{E}_h\|_{\mathcal{T}_h}$		$\|\mathbf{E}-\mathbf{E}_h\|_{\mathbf{H}^c}$		$\|\mathbf{J}-\mathbf{J}_h\|_{\mathcal{T}_h}$		$\|\mathbf{J}-\mathbf{J}_h\|_{\mathbf{H}^d}$	
p	n	Error	Order	Error	Order	Error	Order	Error	Order
1	8	3.09e-2	–	4.21e-1	–	5.02e-2	–	8.67e-1	–
	16	7.21e-3	2.10	2.06e-1	1.02	1.15e-2	2.13	4.22e-1	1.04
	32	1.77e-3	2.03	1.03e-1	1.00	2.79e-3	2.04	2.09e-1	1.02
	64	4.38e-4	2.01	5.11e-2	1.00	6.89e-4	2.02	1.04e-1	1.01
	128	1.09e-4	2.00	2.55e-2	1.00	1.72e-4	2.00	5.18e-2	1.00
2	8	1.22e-3	–	2.27e-2	–	1.80e-3	–	4.49e-2	–
	16	1.47e-4	3.05	5.55e-3	2.03	2.17e-4	3.05	1.11e-2	2.02
	32	1.81e-5	3.02	1.37e-3	2.01	2.68e-5	3.02	2.75e-3	2.01
	64	2.25e-6	3.01	3.42e-4	2.01	3.33e-6	3.01	6.86e-4	2.00
	128	2.81e-7	3.00	8.53e-5	2.00	4.15e-7	3.00	1.71e-4	2.00
3	8	2.91e-5	–	7.93e-4	–	4.46e-5	–	1.64e-3	–
	16	1.79e-6	4.02	9.80e-5	3.02	2.73e-6	4.03	2.02e-4	3.02
	32	1.11e-7	4.01	1.22e-5	3.01	1.69e-7	4.01	2.51e-5	3.01
	64	6.94e-9	4.00	1.52e-6	3.00	1.05e-8	4.01	3.13e-6	3.00
	128	4.33e-10	4.00	1.90e-7	3.00	6.58e-10	4.00	3.90e-7	3.00

Table 3.1: History of convergence of the HDG method for the hydrodynamic model.

$$\begin{aligned}
\|\mathbf{E}-\mathbf{E}_h\|^2_{\mathcal{T}_h} &= \sum_{T\in\mathcal{T}_h}\int_T \|\mathbf{E}-\mathbf{E}_h\|^2, \\
\|\mathbf{J}-\mathbf{J}_h\|^2_{\mathcal{T}_h} &= \sum_{T\in\mathcal{T}_h}\int_T \|\mathbf{J}-\mathbf{J}_h\|^2, \\
\|\mathbf{E}-\mathbf{E}_h\|^2_{\mathbf{H}^c} &= \sum_{T\in\mathcal{T}_h}\int_T \left(\|\mathbf{E}-\mathbf{E}_h\|^2 + \|\nabla\times\mathbf{E}-\nabla\times\mathbf{E}_h\|^2\right), \\
\|\mathbf{J}-\mathbf{J}_h\|^2_{\mathbf{H}^d} &= \sum_{T\in\mathcal{T}_h}\int_T \left(\|\mathbf{J}-\mathbf{J}_h\|^2 + \|\nabla\cdot\mathbf{J}-\nabla\cdot\mathbf{J}_h\|^2\right),
\end{aligned}$$

and set both stabilization parameters according to the values proposed above, that is $\tau_t = \tau_n = \sqrt{2}$. We analyze the convergence of the errors on a sequence of structured triangular meshes with $n^2/2$ elements. We consider polynomials of degree $p = 1$, 2 and 3 to represent the approximate solution, and present the results in Table 3.1. We observe that the HDG

method for the hydrodynamic model converges at the optimal rate of $\mathcal{O}(h^{p+1})$ in the L^2 norm and at the rate of $\mathcal{O}(h^p)$ in the $\mathbf{H}^{\mathrm{div}}(\mathcal{T}_h)$ and $\mathbf{H}^{\mathrm{curl}}(\mathcal{T}_h)$ norms.

3.3.4 Numerical results

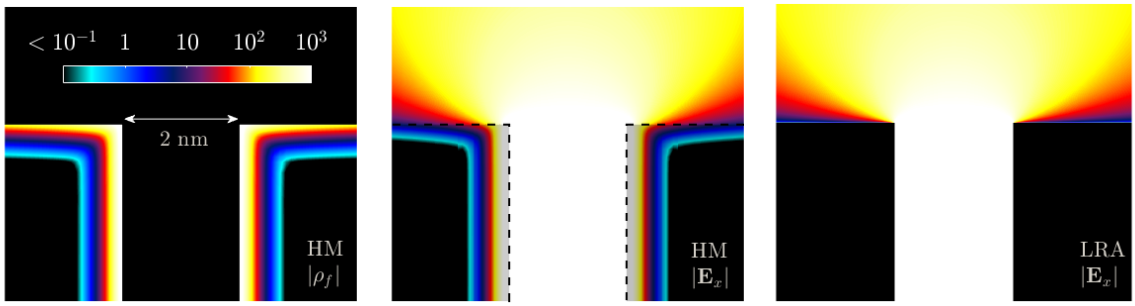
In this section, we present three numerical experiments to assess the performance of the HDG method for the hydrodynamic model.

Nanoslit

For the first example, we revisit the plasmonic nanoslit introduced in Section 3.2.1, and investigate the effect of solving the HM in the computational region comprised by the gold. Including the hydrodynamic currents entails a significant computational burden, since it requires having sufficient spatial resolution to capture the sub-wavelength features that occur in the metal. We focus on very low THz frequencies, for which the longitudinal modes decay exponentially, therefore the main challenge is to capture the profile of induced free charge density at the surface of the metal. The geometric and material parameters are identical to those in Section 3.2.1, and the Fermi velocity for gold is given by $\overline{v}_F = 1.39 \cdot 10^6$ m/s [11].

We employ the same computational domain as for the LRA case, but refining the metallic volume at both sides of the slit, totalling 24K quadrangular cubic elements to accurately capture the complicated features that develop in the HM. In addition to the increased number of discretization elements, accounting for the hydrodynamic equations requires solving for the auxiliary hybrid variable $\widehat{U}_h$. In 2d, this results in doubling the degrees of freedom in the metal, which for the present nanoslit example increases the size of the linear system by more than 20% with respect to the same mesh for the LRA.

The induced free charge density $|\rho_f|$ is shown in Fig. 3-7a for the 2 nm gap at 0.3 THz. To better appreciate the extreme field localization that occurs at the surface of the metal, we focus on the upper tips of the slit and examine the decay away from the tip in logarithmic scale. The charge density experiences an extreme gradient of more than four orders of magnitude within a distance of approximately 1 nm, forming a boundary layer-type structure. Furthermore, we show the enhancement magnitude $|\mathbf{E}_x|$ for the same region in Fig. 3-7b,



(a) Induced free charge density (b) HM field enhancement, dashed metal boundaries (c) LRA field enhancement

Figure 3-7: Upper surface of 2 nm slit illuminated at 0.3 THz.

where we observe that the incident electric field does propagate through gold. Conversely, the enhancement field for the LRA is confined to the aperture, see Fig. 3-7c, since it is not able to displace the electrons in the metal due to the infinite charge density localized at the gold surface. These two solution patterns illustrate the interaction of the electromagnetic incident field with the metallic boundaries, a key difference between LRA and HM.

Plasmonic nanowire dimer

The second example is the plasmonic nanowire dimer, a structure that consists of two infinitely long metallic cylinders in free space that are separated by a small distance, see Fig. 3-8a. When the nanowires are illuminated with a plane wave polarized along the axis of inter-cylinder separation, they are able to support localized plasmon resonances within the gap region, leading to strong field enhancements [185]. The plasmonic dimer has been studied extensively from both the theoretical and the experimental perspective [96], and numerical calculations have been performed for both the LRA [116, 130] and the HM [75, 250].

The plasmonic dimer is interesting because, for gold and silver structures the LRA predicts exponentially increasing field enhancements as the gap vanishes [75, 219], up until the wires touch. This unphysical behavior is corrected by the HM, which incorporates nonlocal interactions between electrons. Moreover, both models predict a different amount of frequency redshift and field enhancement as the gap shrinks.

For this example, we study the differences between both models for dimers consisting of gold supercircles of radius $R = 25$, as the gap varies between 5-25 Å. The supercircle is a closed

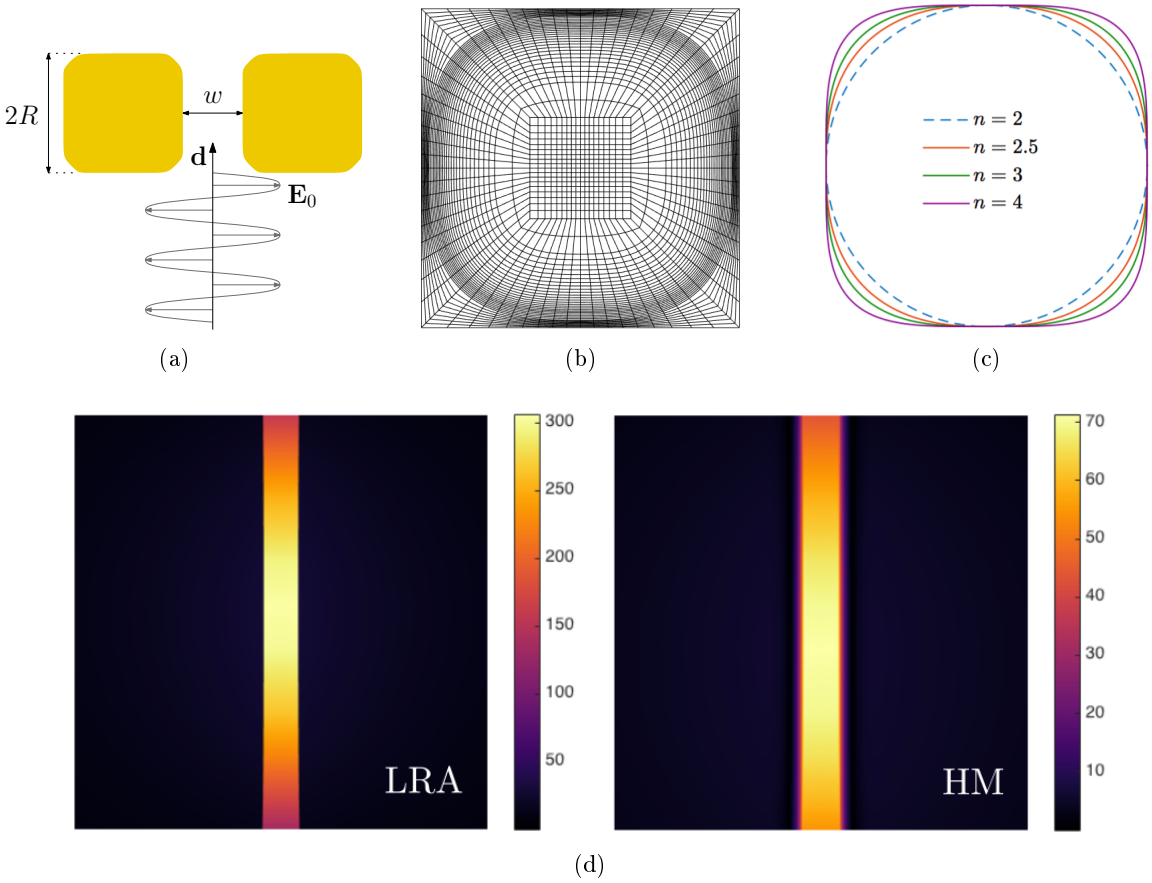


Figure 3-8: (a) Schematic of incident plane wave, polarization and dimensions of gold dimer. (b) Detail of nanowire mesh for $n = 3$. (c) Supercircles for several n. (d) Detail of $|\mathbf{E}_x|$ field for $w = 5$ Å and $n = 3$ for energy $\hbar\overline{\omega} = 2.63$ eV at gap region for LRA (left) and HM (right), which captures metal penetration.

curve with equation

$$|x|^n + |y|^n = R^n, \quad n > 0$$

and we will focus on the case $2 < n \leq 4$, which produces curves between the circle and the square with rounded edges known as squircle, since the circular case has already been extensively characterized [212, 250]. This will allow us to investigate the effect of proximity on the plasmonic response, since increasing n leads to dimers with greater spatial interaction, see Fig. 3-8c. The values for the optical constants read $\hbar\overline{\omega}_p = 9.02$ eV, $\hbar\overline{\gamma} = 0.071$ and $\varepsilon_\infty = 1$ are taken from [119], whereas the Fermi velocity for gold is given by $\overline{v}_F = 1.39 \cdot 10^6$ m/s [11].

For the numerical computations, we use a curved anisotropic mesh of cubic quadrangular elements that is fine in the surface of the dimers and the gap region and coarsens away from

the dimer. In Fig. 3-8b, we show a detail of a single nanowire discretized for a separation of $w = 25$ Å, consisting of 3K elements. Since we want to explore a broad range of frequencies, the discretization away from the gold nanowires is extended and adapated to ensure sufficient spatial resolution is available to accurately simulate the EM propagation. The quantities of interest are the field enhancement π measured in the gap region and the extinction cross section σ_{ext} measured on an arbitrary surface surrounding the dimer.

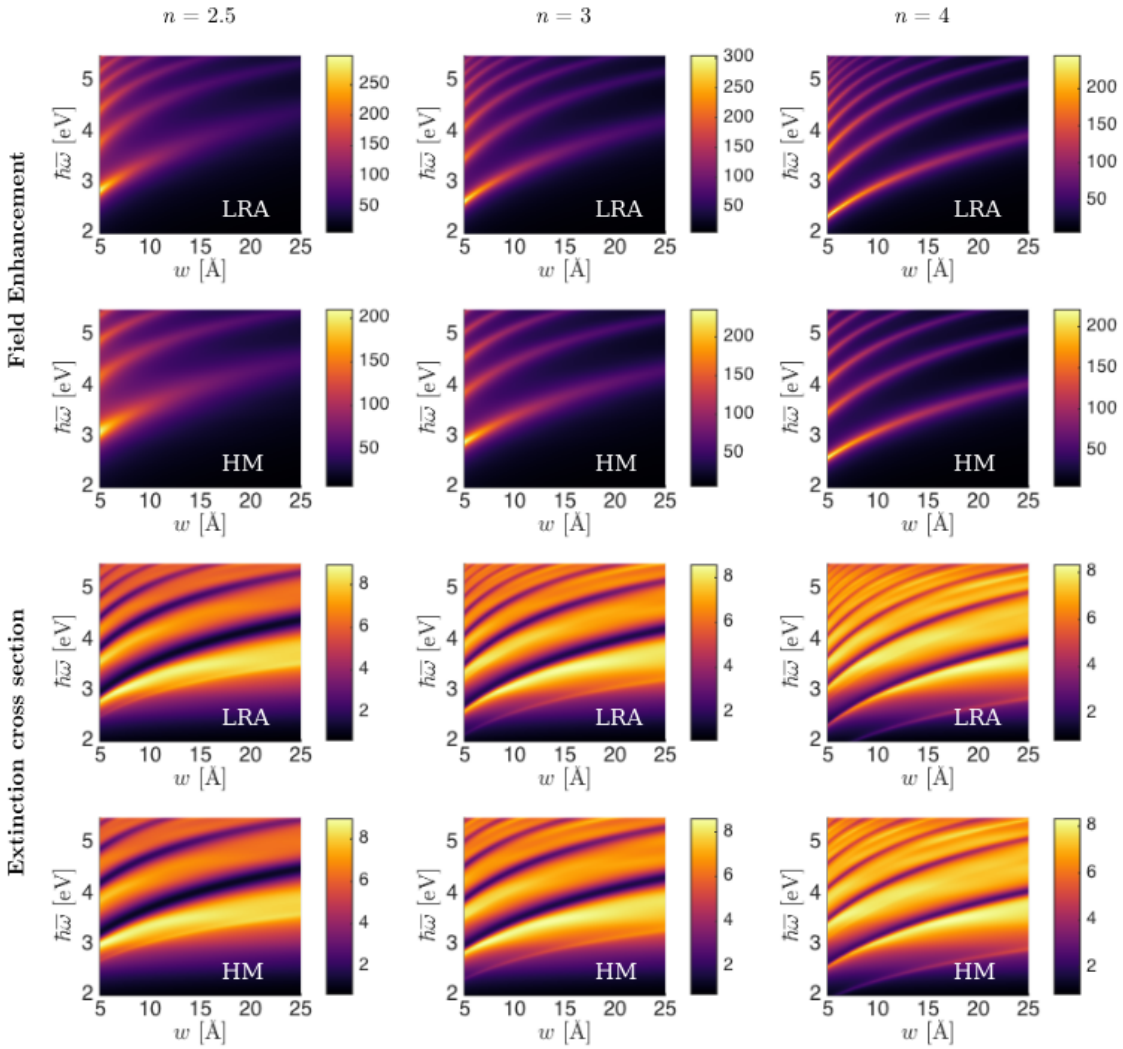


Figure 3-9: Extinction cross section and field enhancement diagram as a function of gap size and energy for supercircle dimers with $n = 2.5, 3$ and 4.

Firstly, to gain insight on the physics we compare the response of the dimer for both models at the resonant photon energy $\hbar\overline{\omega} = 2.63$ eV for $n = 3$ in Fig. 3-8d. The main difference between the LRA and the HM arises at the surface of the nanowire. The Drude model characterizes the metal as hard wall, thus no penetration of the electromagnetic field is

allowed in the metal. Conversely, modeling the nonlocal behavior of the electrons with the hydrodynamic current captures the interaction of the incident field with the electrons. More specifically, the plasmon resonance spreads the electron density distribution at the metal interface, thus the gap is effectively enlarged. As a consequence, the HM predicts lower field enhancements than the LRA.

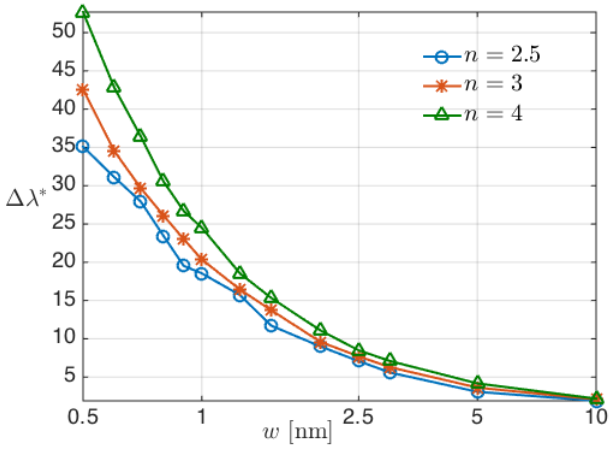


Figure 3-10: Resonance blueshift as a function of dimer separation w in nm for supercircles with $n = 2.5$, 3 and 4.

We now analyze the effect of nonlocality as a function of the gap size, computing the spectra of field enhancement and extinction cross section for different gap sizes and for different geometries of the nanowires. In Fig. 3-9 we collect the results for $n = 2.5\,, 3$ and 4 for both models. Firstly, we find that increasing the geometry parameter n leads to a higher density of resonances within the same energy interval, since the nanowires interact much more closely for the same gap width. In addition, the resonances are narrower and exhibit progressively less overlap. As far as nonlocality is concerned, the main feature that can be observed is that all the plasmon resonances are blueshifted due to the nonlocal pressure term, together with a reduction in the maximum π achieved in the gap, which is more severly affected than σ_{ext}. The nonlocal effects are more predominant as the distance between the wires enters the subnanometer regime, thus stressing the importance of accounting for the hydrodynamic current if accurate calculations are sought. Finally, note that the effect of nonlocality is even more drastic at a fixed frequency, since it can happen that the field enhancement peaks for the LRA while the HM predicts a valley, see for instance the response at 3 eV for $n = 4$.

To better quantify the spectral shift caused by nonlocality, we show in Fig. 3-10 the blueshift

$\Delta\lambda^* = \lambda^*_{LRA} - \lambda^*_{HM}$ in nm, that is the deviation between wavelengths for both models at the first resonance, as a function of the separation between nanowires for the geometries $n = 2.5, 3$ and 4. As anticipated by the results in Fig. 3-9, the blueshift introduced by the nonlocal model is larger for supercircles with greater n, as the nanowires interact along a more extended area. In all cases, the effects of nonlocality diminish as the gap size increases, and eventually become negligible for distances greater than 10 nm, as predicted by the hydrodynamic theory.

Periodic annular array

Lastly, we analyze the impact of nonlocality in the 3d annular structure introduced in Fig. 3-5a. Similarly as in the case for the nanoslit, we refine the discretization for the metal to ensure we capture the tight localization of the charge density at the metal-dielectric interface. This refinement results in 200 extra elements in the metal, which translates into an increase of 26% in global degrees of freedom. The reduced number of globally coupled degrees of freedom of the HDG method is crucial, since instead of solving for ten (resp. six) volume unknowns in the metal (resp. dielectric), we are left with a linear system involving only three (resp. two) face unknowns in the metal (resp. dielectric). The linear system is of dimension 345K, and MATLAB's sparse direct solver is again utilized to obtain the solution. Nonetheless, the computational cost and memory usage when considering the HM increases significantly, since we not only have more degrees of freedom, but also the extremely localized features of the solution field in the metal need to be resolved.

In Fig. 3-11a, we inspect the induced free charge density $|\rho_f|$ for several angular slices as before, where we observe for $\alpha = 0°$ the same boundary-layer pattern as in the nanoslit, with a maximum value at the interface and a decay of five orders of magnitude just a few nanometers away from the aperture. These two features gradually decrease as we move from the y-constant symmetry plane to the x-constant symmetry plane. Indeed, for $\alpha = 75°$ the charge density profile is almost constant in the interior of the metal.

Finally, we study the effect of nonlocality on the resonance that occurs at 0.6 THz. The inclusion of the hydrodynamic current results in a relative blueshift of around 2%, along with a 1.7% decrease in the field enhancement, shown in Fig. 3-11b. This behavior occurs as a

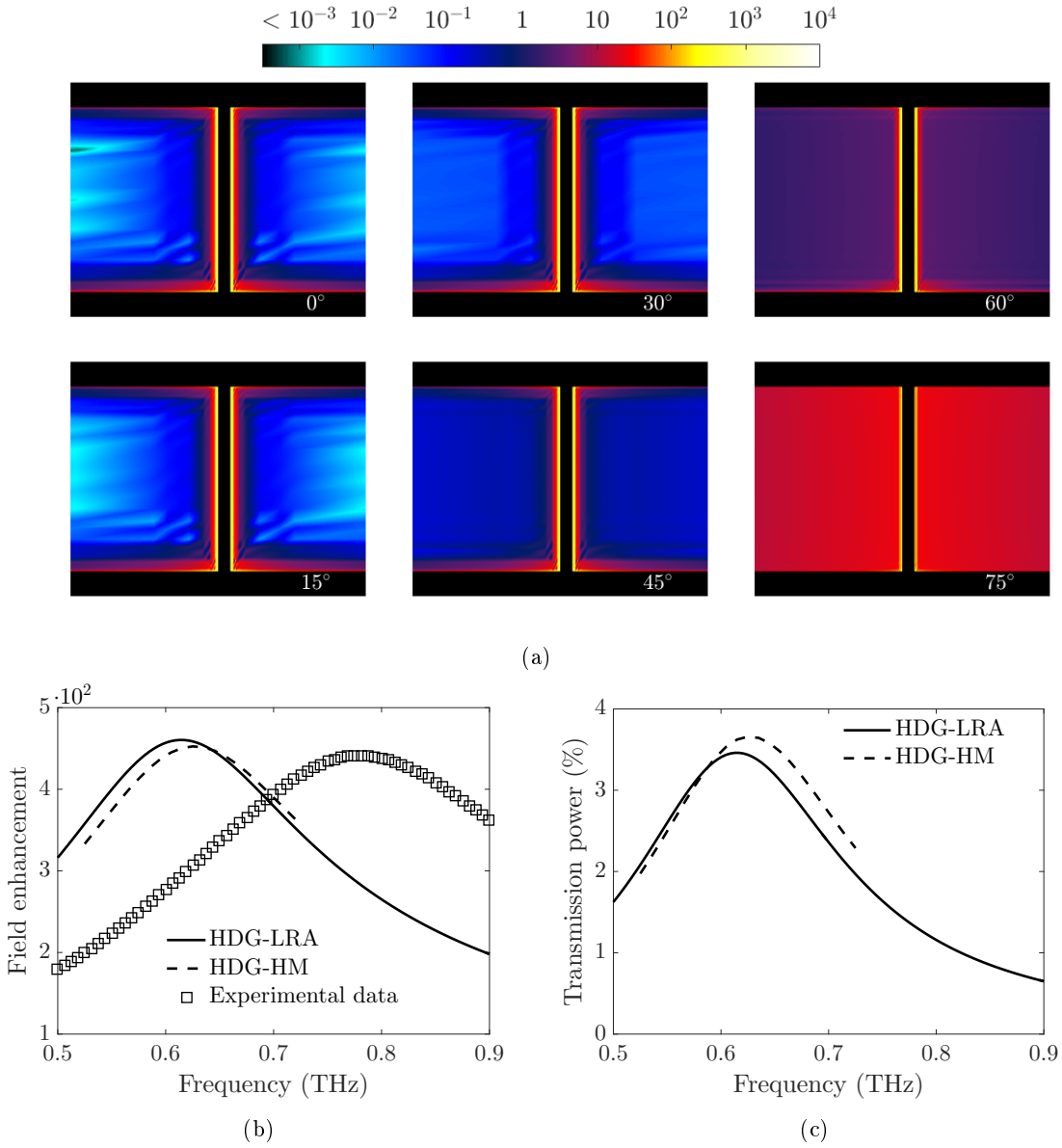


Figure 3-11: (a) Slices in the angular direction of solution field $|\rho_f|$ at the aperture for 0.6 THz, shown in logarithmic scale. (b) Field enhancement curve for LRA and HM compared with experimental data. (c) Transmission power curve for LRA and HM.

consequence of the penetration of the EM field into the metal, thus relaxing the infinitely concentrated charge density modeled by the LRA. The smoothed charge density profile effectively implies a widening of the aperture, which results in a: (1) resonance blueshift, that is, a shorter wavelength is required to excite the plasmon resonance; (2) a decrease in field enhancement, since there is less confinement; and (3) a 6% increase in transmitted power as the EM wave can propagate through a greater aperture space, see Fig. 3-11c.

Nonetheless, incorporating nonlocality does not bridge the gap between simulations and experimental data. Moreover, it is unlikely that considering quantum effects would produce further blueshifts, since for 10 nm separation between gold surfaces electron tunnelling is negligible. In Chapter 5, we extend the study of discrepancies between experiments and simulations to include material properties and geometric features.

3.4 Concluding remarks

In this chapter, we have presented a set of hybridizable discontinuous Galerkin methods to simulate the propagation of electromagnetic waves for metal-dielectric media at the nanoscale. Simulation of plasmonic phenomena is inherently complex due to the enormous disparity in length scales and the extreme localization of electromagnetic fields that can be observed as a consequence of the collective excitation of electrons. The HDG method for Maxwell's equations, and the extension to account for the hydrodynamic model for metals are powerful methods to tackle the numerical simulation of plasmonic structures, due to its ability to handle complex geometries through anisotropic unstructured meshes, the efficient treatment of material interfaces and the possibility of solving reduced linear systems that only involve the degrees of freedom at the faces of the discretization. This aspect can be of significant advantage since iterative and preconditioning techniques for electromagnetic wave propagation problems is still an active area of research.

These numerical techniques have been applied to the forward simulation of a 2d nanoslit, a 2d nanowire dimer and a 3d periodic annular structure, for which high-fidelity solvers are required to properly capture the full electromagnetic field.

Chapter 4

Reduced order modeling for plasmonics

Reduced order models (ROM) can be used for simulating large-scale systems when multiple evaluations of the forward problem for different values of the input parameters are required. The objective is to develop a surrogate model that produces accurate solutions and can be evaluated in real-time, thus alleviating the computational cost of multiple full model solutions. The basic idea behind these methods relies on projecting the high-dimensional state space onto a much lower dimension state space, thus creating a reduced model with a reduced state space.

In this chapter, we describe the implementation of ROM for the time-harmonic Maxwell's equations in (3.1) and (3.2) using HDG and proper orthogonal decomposition techniques. The reduction of material parameters is pursued with a new weak formulation for the HDG method that is optimal in terms of degrees of freedom. In addition, the reduction of geometric parameters is achieved by transforming the governing equations using a deformation mapping and a fixed reference domain.

4.1 HDG-POD method for Maxwell's equations

In this section, we develop a reduced order model for the HDG discretization of time-harmonic Maxwell's equations introduced in (3.6) based on the POD. For the construction of the ROM, we first need to identify the relevant parameters of interest. For the regular Maxwell's equations we are interested in the frequency ω and the material properties encap-

sulated in ε. Note that, for the case of metals, the dispersive permittivity is given by the Drude expression (2.7), thus the plasma frequency ω_p and the collision rate γ are treated as parameters. All in all, and since multiple materials may be present in the simulation, Q shall denote the total number of parameters.

A feature of the HDG discretization is that the affine parametric dependence present in (3.6) is lost in the solution process. Indeed, in order to eliminate the local degrees of freedom $\mathbf{V}_h$, $\mathbf{E}_h$ and arrive to a system involving only the global degrees of freedom $\widehat{\mathbf{E}}_h$, we need to substitute the first two equations (3.6a) and (3.6b) into (3.6c), rendering a matrix system that is *nonaffine* in ω, ε. To avoid this inconvenient situation, we propose a new HDG formulation for Maxwell's equations that naturally enables the application of POD techniques.

4.1.1 A new weak formulation for the HDG method

We begin by deriving a new weak formulation for Maxwell's equations, inspired in the weak formulation introduced in [261] for the Helmholtz equation. We first introduce two lifting operators $\mathbf{b} : \boldsymbol{W}_h \to \boldsymbol{W}_h$ and $\mathbf{c} : \boldsymbol{M}_h \to \boldsymbol{W}_h$ defined as:

$$\begin{aligned} (\mathbf{b}(\boldsymbol{\xi}), \boldsymbol{\kappa})_{\mathcal{T}_h} &= (\boldsymbol{\xi}, \nabla \times \boldsymbol{\kappa})_{\mathcal{T}_h}, \qquad \forall \boldsymbol{\kappa} \in \boldsymbol{W}_h \\ (\mathbf{c}(\boldsymbol{\mu}), \boldsymbol{\kappa})_{\mathcal{T}_h} &= \langle \boldsymbol{\xi}, \boldsymbol{\kappa} \times \mathbf{n} \rangle_{\mathcal{T}_h}, \qquad \forall \boldsymbol{\kappa} \in \boldsymbol{W}_h \end{aligned} \tag{4.1}$$

It thus follows from (3.6a) and (4.1) that we can express $\mathbf{V}_h$ as a function of $\mathbf{E}_h$, $\widehat{\mathbf{E}}_h$ as

$$\mathbf{V}_h = \mathbf{b}(\mathbf{E}_h) + \mathbf{c}(\widehat{\mathbf{E}}_h)\,. \tag{4.2}$$

If we substitute (4.2) into (3.6b) and (3.6c) we obtain the following weak formulation: find $(\mathbf{E}_h, \widehat{\mathbf{E}}_h) \in \boldsymbol{W}_h \times \boldsymbol{M}_h(\mathbf{E}_D)$ such that

$$\begin{aligned} \big(\nabla \times (\mathbf{b}(\mathbf{E}_h) + \mathbf{c}(\widehat{\mathbf{E}}_h)), \boldsymbol{\xi}\big)_{\mathcal{T}_h} + \langle \tau_t [\mathbf{E}_h - \widehat{\mathbf{E}}_h], \mathbf{n} \times \boldsymbol{\xi} \times \mathbf{n} \rangle_{\partial \mathcal{T}_h} - \omega^2 (\varepsilon \mathbf{E}_h, \boldsymbol{\xi})_{\mathcal{T}_h} = 0, \\ -\langle \mathbf{n} \times (\mathbf{b}(\mathbf{E}_h) + \mathbf{c}(\widehat{\mathbf{E}}_h)) + \tau_t \mathbf{E}_h, \boldsymbol{\mu} \rangle_{\partial \mathcal{T}_h \setminus \partial \Omega_D} + \langle \widetilde{\tau}_t \widehat{\mathbf{E}}_h, \boldsymbol{\mu} \rangle_{\partial \mathcal{T}_h} = \langle \mathbf{F}(\omega, \varepsilon), \boldsymbol{\mu} \rangle_{\partial \Omega}. \end{aligned}$$

for all $(\boldsymbol{\xi}, \boldsymbol{\mu}) \in \boldsymbol{W}_h \times \boldsymbol{M}_h(\mathbf{0})$, where the boundary flux $\mathbf{F}$ is given by (3.4). Then, we set the $\mathcal{N}^{\Upsilon}$-dimensional approximation space to be $\boldsymbol{W}_h^{\Upsilon} := \boldsymbol{W}_h \times \boldsymbol{M}_h(\mathbf{E}_D)$, $\mathbf{E}_h^{\Upsilon} := (\mathbf{E}_h, \widehat{\mathbf{E}}_h)$

and $\boldsymbol{\xi}^{\Upsilon} := (\boldsymbol{\xi}, \boldsymbol{\mu})$ such that the weak formulation for Maxwell's equation can be compactly written, for $\mathbf{E}_h^{\Upsilon} \in \boldsymbol{W}_h^{\Upsilon}$, as

$$\mathcal{A}^{\Upsilon}(\mathbf{E}_h^{\Upsilon}, \boldsymbol{\xi}^{\Upsilon}; (\omega, \varepsilon)) = \mathcal{F}^{\Upsilon}(\boldsymbol{\xi}^{\Upsilon}; (\omega, \varepsilon)), \qquad \forall \boldsymbol{\xi}^{\Upsilon} \in \boldsymbol{W}_h \times \boldsymbol{M}_h(\mathbf{0}) \tag{4.3}$$

where the bilinear form $\mathcal{A}$ and the linear functional $\mathcal{F}^{\Upsilon}$ are given by

$$\begin{aligned}
\mathcal{A}^{\Upsilon}(\mathbf{u}^{\Upsilon}, \boldsymbol{\xi}^{\Upsilon}; (\omega, \varepsilon)) =& \big(\nabla \times (\mathbf{b}(\mathbf{u}) + \mathbf{c}(\boldsymbol{\nu})), \boldsymbol{\xi}\big)_{\mathcal{T}_h} + \langle \tau_t[\mathbf{u} - \boldsymbol{\nu}], \mathbf{n} \times \boldsymbol{\xi} \times \mathbf{n} \rangle_{\partial \mathcal{T}_h} - \omega^2 (\varepsilon \mathbf{u}, \boldsymbol{\xi})_{\mathcal{T}_h} \\
& - \langle \mathbf{n} \times (\mathbf{b}(\mathbf{u}) + \mathbf{c}(\boldsymbol{\nu})) + \tau_t \mathbf{u}, \boldsymbol{\mu} \rangle_{\partial \mathcal{T}_h \setminus \partial \Omega_D} + \langle \widetilde{\tau}_t \boldsymbol{\nu}, \boldsymbol{\mu} \rangle_{\partial \mathcal{T}_h}, \\
\mathcal{F}^{\Upsilon}(\boldsymbol{\xi}^{\Upsilon}; (\omega, \varepsilon)) =& \langle \mathbf{F}(\omega, \varepsilon), \boldsymbol{\mu} \rangle_{\partial \Omega},
\end{aligned}$$

for all $\mathbf{u}^{\Upsilon} = (\mathbf{u}, \boldsymbol{\nu}) \in \boldsymbol{W}_h^{\Upsilon}$. The key aspect of this weak formulation is that the new bilinear and linear forms are affine in (ω, ε). Furthermore, this new formulation is optimal in terms of degrees of freedom, since we no longer account for $\mathbf{V}_h$.

In addition, we shall define an inner product for the high-dimensional approximation space $\boldsymbol{W}_h^{\Upsilon}$ as

$$(\mathbf{u}^{\Upsilon}, \boldsymbol{\xi}^{\Upsilon})_{\boldsymbol{W}^{\Upsilon}} = (\mathbf{u}, \boldsymbol{\xi})_{\mathcal{T}_h} + \langle \boldsymbol{\nu}, \boldsymbol{\mu} \rangle_{\partial \mathcal{T}_h}. \tag{4.4}$$

The inner product $(\cdot, \cdot)_{\boldsymbol{W}^{\Upsilon}}$ defines an induced norm $\|\boldsymbol{\xi}^{\Upsilon}\|_{\boldsymbol{W}^{\Upsilon}} = \sqrt{(\boldsymbol{\xi}^{\Upsilon}, \boldsymbol{\xi}^{\Upsilon})_{\boldsymbol{W}^{\Upsilon}}}$, and allows us to consistently define the inner products between elements of $\boldsymbol{W}_h^{\Upsilon}$.

We note that, even though the formulation in (4.3) involves $\mathbf{E}_h$, these degrees of freedom are never computed when solving the full system (4.3). Instead, we invoke once more discontinuity of the approximation spaces and express $\mathbf{E}_h$ in terms of $\widehat{\mathbf{E}}_h$, hence solving linear systems involving only the global degrees of freedom $\widehat{\mathbf{E}}_h$.

4.1.2 POD formulation

We assume that we are given a collection of solutions, or snapshots, $\mathbf{E}_h^{\Upsilon} \in \boldsymbol{W}_h^{\Upsilon}$, computed at J selected parameter values, stored by columns in a matrix $[\mathbf{E}_h^{\Upsilon}(\omega_1, \varepsilon_1), \ldots, \mathbf{E}_h^{\Upsilon}(\omega_J, \varepsilon_J)]$, along with the matrix arising from the inner product defined in (4.4). We then apply the POD algorithm [265] and obtain a set of $N_{\max}$ orthonormalized basis functions $\boldsymbol{\zeta}_n^{\Upsilon} \in \boldsymbol{W}_h^{\Upsilon}$, $1 \leq n \leq N_{\max}$ such that $(\boldsymbol{\zeta}_n^{\Upsilon}, \boldsymbol{\zeta}_{n'}^{\Upsilon})_{\boldsymbol{W}^{\Upsilon}} = \delta_{nn'}$ for $1 \leq n, n' \leq N_{\max}$. 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 [265] for more details.

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

$$\boldsymbol{W}_N^\Upsilon = \text{span}\{\boldsymbol{\zeta}_n^\Upsilon,\, 1 \leq n \leq N\}, \quad N = 1, \ldots, N_{\max}.$$

The HDG-POD method results from applying a Galerkin projection (4.3) using the POD space, that is: for a given pair (ω, ε), find an approximation $\mathbf{E}_N^\Upsilon(\omega, \varepsilon) \in \boldsymbol{W}_N^\Upsilon$ satisfying

$$\mathcal{A}^\Upsilon(\mathbf{E}_N^\Upsilon, \boldsymbol{\xi}^\Upsilon\,; (\omega, \varepsilon)) = \mathcal{F}^\Upsilon(\boldsymbol{\xi}^\Upsilon\,; (\omega, \varepsilon)), \qquad \forall \boldsymbol{\xi}^\Upsilon \in \boldsymbol{W}_N^\Upsilon\,, \tag{4.5}$$

where we solve for the trial coefficients $\{\lambda_n^\Upsilon\}_{n=1}^N$ and $\mathbf{E}_N^\Upsilon = \sum_{n=1}^N \lambda_n^\Upsilon \boldsymbol{\zeta}_n^\Upsilon$. Additionally, we can recover the approximate field $\mathbf{V}_N$ using expression (4.2) with $\mathbf{E}_N$ and $\widehat{\mathbf{E}}_N$. In principle, we expect the cost of (4.5) to be much smaller than that of (4.3), since $N \ll \mathcal{N}^\Upsilon$.

4.1.3 Computational strategy

The linearity and 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 the computation of J snapshots, the POD compression that produces the orthonormalized snapshots $\boldsymbol{\zeta}_n^\Upsilon, 1 \leq n \leq N$ associated with the HDG approximation space at the selected parameter values and the formation and storage of several parameter-independent small matrices and vectors. The online stage – parameter dependent, performed multiple times – evaluates the trial coefficients $\{\lambda_n^\Upsilon(\omega, \varepsilon)\}_{n=1}^N$ for any new (ω, ε) with complexity $\mathcal{O}\left(N^3 + QN^2\right)$, independent of the dimension $\mathcal{N}^\Upsilon$ of the HDG approximation space.

After the trial coefficients have been computed, recovering the approximate field variables $\mathbf{V}_N$, $\mathbf{E}_N$ still requires querying the POD basis with complexity $\mathcal{O}\left(4N\mathcal{N}^\Upsilon\right)$. In the RB community this obstacle is avoided by never evaluating the field variables; if the QoI is *linear* in the field variables, the offline-online strategy enables dropping the $\mathcal{N}^\Upsilon$ dependence in the online stage, see [181, 206]. Unfortunately, the quantities of interest in plasmonics simulations, such as the field enhancement or the optical intensity introduced in Section 2.3.1, will

in general be nonlinear in the field variables, thus the complexity for the online stage will be $\mathcal{N}^{\Upsilon}$-dependent. Despite this shortcoming, if the QoI 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.

In summary, the implications of the above strategy are twofold: first, if N and Q are small and the QoI 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.

4.1.4 Numerical results

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 employ the HDG-POD technique on a realistic 3d example and treat the material properties as exploration parameters. The main objective is to show that, with few full model evaluations, we can construct a reduced order model that is capable of predicting, in real-time, the response from the plasmonic device for a range of material properties and frequencies.

Firstly, we shall define the parameters and their corresponding intervals of variability. We are interested in studying the field enhancement π on the aperture volume of the annular structure introduced in Fig. 3-5a, with frequencies ranging from 0.5 to 0.9 THz. The only difference here is that we consider a sapphire substrate, which is also transparent at low THz frequencies [91]. In Table 4.1, we summarize the values used for the parameters, along with the references wherefrom data was extracted. For the optical constants of gold, we have considered the range of variation given by [197, 208], even though it corresponds to a different frequency regime.

Secondly, the matrix of snapshots is obtained by sampling the 5-dimensional parameter space (ω + 4 material parameters) and computing the solution $\mathbf{E}_h^{\Upsilon}$ at 350 selected parameter values. We use quasi random sequences [31, 182], in particular the Sobol sequence [240], to

Parameter	Values	Frequency range (μm)	References
$\hbar\overline{\omega}_p$ [eV]	9.02 ± 0.18	< 12	[193, 197, 208]
$\hbar\overline{\gamma}$ [eV]	0.02678 ± 0.007	< 12	[193, 197, 208]
$\varepsilon_{Al_2O_3}$	5 - 6	thickness-dependent	[92]
n_{sapphire}	3.07 ± 0.006	< 2	[91]

Table 4.1: Variability ranges for parameters, interval of validity and references for 10 nm alumina gap in low THz frequencies.

achieve a more uniform sampling in the high-dimensional space, although techniques such as Latin Hypercube sampling [143, 244], sparse grids [239] or plain random number generation are also valid strategies.

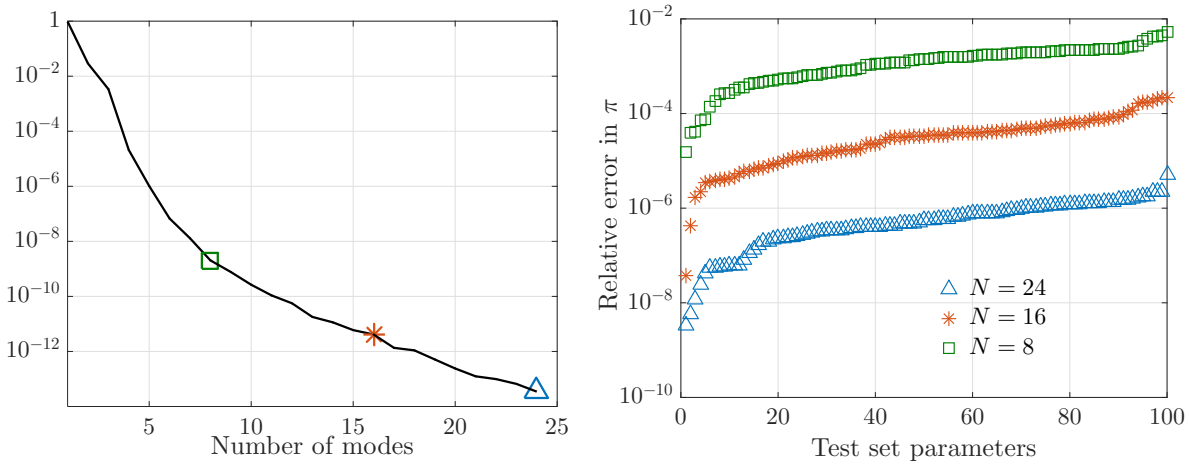


Figure 4-1: (a) Decay of normalized singular values vs number of POD modes N. (b) Relative errors in field enhancement for test set parameters, evaluated for multiple POD sizes.

In Fig. 4-1a we show the decay of the normalized singular values of the snapshot matrix up to $N = 24$. In order to gauge the accuracy of the POD basis, we compute the HDG solution for 100 random parameter values –referred to as test set– and evaluate the relative error committed by different fidelities of the reduced order model on the test set of examples. The results are collected in Fig. 4-1b, where for each POD basis size N we sort the relative errors. It should be observed that with 16 POD modes the relative error for all testing examples is well below 0.1%, which is generally deemed sufficient for most engineering applications.

The efficiency of the POD basis is assessed timing its online performance, more specifically the assembly, solution of the linear system and recovery of the local variables from the trial coefficients $\{\lambda_n^\Upsilon\}_{n=1}^N$. Time estimates are obtained averaging the wall time for 100 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 4.2 in milliseconds. The advantages of the reduced order model are apparent 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. 4-1b. As anticipated, the bulk of the computational cost to evaluate the POD basis is devoted to recovering the local variables required to evaluate the QoI, since it involves operating with the high-dimensional POD basis functions. This approach is particularly beneficial for large problems with parametric variability, where

Model	Assembly (ms)	Linear system (ms)	Local variables (ms)
$N = 8$	0.12	0.11	5.33
$N = 16$	0.16	0.12	5.91
$N = 24$	0.21	0.18	6.51
HDG	9.4e4	2.72e5	850

Table 4.2: Computational wall time in milliseconds of HDG/HDG-POD for time-harmonic Maxwell's equations.

the model must produce accurate solutions in real-time for multiple queries.

4.2 HDG-POD method for Maxwell's equations in the reference domain

The reduced order model constructed above allows us to explore a variety of material configurations at a cost much smaller than that of the full electromagnetic simulation with the HDG method. However, besides material parameters it is often interesting to analyze the electromagnetic response under different geometric configurations. In this section, we extend the HDG-POD method introduced above to include geometric variations in the computational domain. In principle, accounting for geometry variations is nontrivial, since we want to avoid redescritizing the computational domain for each new deformation. Instead, we develop a reference domain formulation for the time-harmonic Maxwell's equations in which a reference domain is mapped to the physical domain by means of a parametrized diffeomorphism $\mathfrak{G}(\cdot, \boldsymbol{\theta})$. More specifically, the main idea is to obtain a reduced order model for Maxwell's equations that comprises not only frequency and material properties, but also the values $\boldsymbol{\theta}$ that parametrize the diffeomorphism.

This approach enables us to solve, for any geometric configuration, a modified version of

Maxwell's equations on a fixed, parameter-independent, reference domain. We now introduce a reference domain formulation for the time-harmonic Maxwell's equations, review its implementation and discuss the application of interpolation techniques to incorporate the geometric parameters into the reduced order model.

4.2.1 Maxwell's equations in reference domain

Let us assume that the physical domain Ω is subject to a deformation mapping that modifies its geometry, parametrized by $\boldsymbol{\theta}$ and defined on a compact set. We want to solve problem (3.1) for many different realizations of $\boldsymbol{\theta}$. In such scenario, it is much more convenient to map the physical domain Ω onto a fixed reference domain $\Omega_{\mathbf{r}}$. Following Persson *et al.* [201], we assume a one-to-one mapping given by a diffeomorphism $\mathfrak{G}$ from the reference domain $\Omega_{\mathbf{r}} \in \mathbb{R}^n$ with coordinates $\mathbf{x}_{\mathbf{r}}$ to the physical domain $\Omega \in \mathbb{R}^n$ with coordinates $\mathbf{x}$. The mapping can be expressed as $\mathbf{x} = \mathfrak{G}(\mathbf{x}_{\mathbf{r}}, \boldsymbol{\theta})$. The mapping deformation gradient 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 scatterer, thus the outer boundaries of the computational domain –where radiation conditions are prescribed– remain unaltered. Starting from $\mathcal{L}$ in (3.1), to obtain the transformed equation on the reference space $\Omega_{\mathbf{r}}$, we integrate on a control volume $v \in \Omega$ and use Stokes' theorem on the curl, namely

$$\begin{aligned}\int_v \nabla \times \mathbf{V} dv = -\int_s \mathbf{V} \times \mathbf{n} ds = -\int_{s_{\mathbf{r}}} g\mathbf{V} \times \left(\mathcal{G}^{-T}\mathbf{n}_{\mathbf{r}}\right) ds_{\mathbf{r}} = -\int_{s_{\mathbf{r}}} g\left(\mathcal{G}^{-T}\left(\mathcal{G}^{T}\mathbf{V}\right)\right) \times \left(\mathcal{G}^{-T}\mathbf{n}_{\mathbf{r}}\right) ds_{\mathbf{r}} \\ = -\int_{s_{\mathbf{r}}} \mathcal{G}\left(\mathcal{G}^{T}\mathbf{V}\right) \times \mathbf{n}_{\mathbf{r}} ds_{\mathbf{r}} = \int_{v_{\mathbf{r}}} \mathcal{G}\nabla_{\mathbf{r}} \times \left(\mathcal{G}^{T}\mathbf{V}\right) dv_{\mathbf{r}}\end{aligned}$$

where we used the identity $Mu \times Mw = (\det M)\, M^{-T}(u \times w)$ and the relations $dv = g dv_{\mathbf{r}}$ and $\mathbf{n} ds = g\mathcal{G}^{-T}\mathbf{n}_{\mathbf{r}} ds_{\mathbf{r}}$. The other term in the equation transforms as

$$\int_v \omega^2 \varepsilon \mathbf{E} dv = \int_V \omega^2 \varepsilon \mathbf{E} g dv_{\mathbf{r}}$$

The boundary conditions can be transformed similarly, for instance

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

Finally, the time-harmonic Maxwell's equations read

$$\begin{aligned} \mathbf{v} - \nabla_\mathbf{r} \times \boldsymbol{G}\mathbf{e} &= 0, \\ \nabla_\mathbf{r} \times \boldsymbol{G}\mathbf{v} - \omega^2 \varepsilon \mathbf{e} &= 0, \end{aligned} \tag{4.6}$$

along with boundary conditions

$$\begin{aligned} &\mathbf{n}_\mathbf{r} \times (\boldsymbol{G}\mathbf{e}) = \mathbf{n}_\mathbf{r} \times (\boldsymbol{G}\mathbf{e}_D), \quad \text{on } \partial\Omega_{\mathbf{r},D}, \\ &\mathbf{n}_\mathbf{r} \times (\boldsymbol{G}\mathbf{v}) = \mathbf{n}_\mathbf{r} \times (\boldsymbol{G}\mathbf{v}_N), \quad \text{on } \partial\Omega_{\mathbf{r},N}, \\ &\mathbf{n}_\mathbf{r} \times (\boldsymbol{G}\mathbf{v}) + i\omega\sqrt{\varepsilon} \mathbf{n}_\mathbf{r} \times (\boldsymbol{G}\mathbf{e}) \times \mathbf{n}_\mathbf{r} = \mathbf{f}^{inc}, \quad \text{on } \partial\Omega_{SM}, \end{aligned}$$

where a generic field $\mathbf{u}$ transforms as $\mathbf{u} = g\mathcal{G}^{-1}\mathbf{U}$ and $\boldsymbol{G} = g^{-1}\mathcal{G}^T\mathcal{G}$. For the reference domain case, we do not need to transform the radiation conditions, since deformation is only prescribed in the vicinity of the scatterer and the Silver-Müller boundary remains unaltered. Thus, we have $\mathbf{V} = \boldsymbol{G}\mathbf{v}$ and $\mathbf{E} = \boldsymbol{G}\mathbf{e}$ on $\partial\Omega_{SM}$, since $\mathfrak{G}\restriction_{\partial\Omega_{SM}} = \mathrm{Id}$.

In order to solve (4.6), we first seek an approximation $(\mathbf{v}_h, \mathbf{e}_h) \in \boldsymbol{W}_h \times \boldsymbol{W}_h$ to the transformed electric and magnetic 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 numerical traces of $\widehat{\mathbf{v}}_h$, 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 (4.6) 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 $(\boldsymbol{\kappa}, \boldsymbol{\xi}, \boldsymbol{\mu}) \in \boldsymbol{W}_h \times \boldsymbol{W}_h \times \boldsymbol{M}_h(\mathbf{0})$

$$\begin{aligned} (\mathbf{v}_h, \boldsymbol{\kappa})_{\mathcal{T}_h} - (\boldsymbol{G}\mathbf{e}_h, \nabla \times \boldsymbol{\kappa})_{\mathcal{T}_h} - \langle \boldsymbol{G}\widehat{\mathbf{e}}_h, \boldsymbol{\kappa} \times \mathbf{n}_\mathbf{r} \rangle_{\partial\mathcal{T}_h} &= 0, \\ (\boldsymbol{G}\mathbf{v}_h, \nabla \times \boldsymbol{\xi})_{\mathcal{T}_h} + \langle \boldsymbol{G}\widehat{\mathbf{v}}_h, \boldsymbol{\xi} \times \mathbf{n}_\mathbf{r} \rangle_{\mathcal{T}_h} - \omega^2 (\varepsilon \mathbf{e}_h, \boldsymbol{\xi})_{\mathcal{T}_h} &= 0, \\ -\langle \mathbf{n}_\mathbf{r} \times \boldsymbol{G}\widehat{\mathbf{v}}_h, \boldsymbol{\mu} \rangle_{\partial\mathcal{T}_h \setminus \partial\Omega_{\mathbf{r},D}} + \langle \boldsymbol{G}\widehat{\mathbf{e}}_h, \boldsymbol{\mu} \rangle_{\partial\Omega_{\mathbf{r},D}} - i\omega \langle \sqrt{\varepsilon} \boldsymbol{G}\widehat{\mathbf{e}}_h, \boldsymbol{\mu} \rangle_{\partial\Omega_{SM}} &= \langle \mathbf{f}, \boldsymbol{\mu} \rangle_{\partial\Omega_\mathbf{r}}, \end{aligned}$$

with

$$\mathbf{f} = \boldsymbol{G}\mathbf{e}_D\restriction_{\partial\Omega_D} - \mathbf{n}\times\boldsymbol{G}\mathbf{v}_N\restriction_{\partial\Omega_N} - \mathbf{f}^{inc}\restriction_{\partial\Omega_{SM}}. \tag{4.7}$$

The tangential component of $\mathbf{v}_h$ is approximated as before

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

Introducing this approximation in the above system, we arrive to the final HDG system for the frequency-domain Maxwell's equations on a reference domain:

$$(\mathbf{v}_h,\boldsymbol{\kappa})_{\mathcal{T}_h} - (\boldsymbol{G}\mathbf{e}_h,\nabla\times\boldsymbol{\kappa})_{\mathcal{T}_h} - \langle\boldsymbol{G}\widehat{\mathbf{e}}_h,\boldsymbol{\kappa}\times\mathbf{n}_{\mathbf{r}}\rangle_{\partial\mathcal{T}_h} = 0, \tag{4.8a}$$

$$\begin{aligned}(\boldsymbol{G}\mathbf{v}_h,\nabla\times\boldsymbol{\xi})_{\mathcal{T}_h} + \langle\boldsymbol{G}\mathbf{v}_h,\boldsymbol{\xi}\times\mathbf{n}_{\mathbf{r}}\rangle_{\mathcal{T}_h} - \omega^2(\varepsilon\mathbf{e}_h,\boldsymbol{\xi})_{\mathcal{T}_h}\\ +\langle\tau_t\boldsymbol{G}(\mathbf{e}_h-\widehat{\mathbf{e}}_h)\times\mathbf{n}_{\mathbf{r}},\boldsymbol{\xi}\times\mathbf{n}_{\mathbf{r}}\rangle_{\mathcal{T}_h} = 0,\end{aligned} \tag{4.8b}$$

$$-\langle\mathbf{n}_{\mathbf{r}}\times\boldsymbol{G}\mathbf{v}_h + \tau_t\boldsymbol{G}\mathbf{e}_h,\boldsymbol{\mu}\rangle_{\partial\mathcal{T}_h\setminus\partial\Omega_{\mathbf{r},D}} + \langle\widetilde{\tau}_t\boldsymbol{G}\widehat{\mathbf{e}}_h,\boldsymbol{\mu}\rangle_{\partial\mathcal{T}_h} = \langle\mathbf{f},\boldsymbol{\mu}\rangle_{\partial\Omega_{\mathbf{r}}}, \tag{4.8c}$$

where the stabilization constant $\widetilde{\tau}_t$ is given in (3.7). Solving the weak formulation above is equivalent to solving (3.6) after deforming the domain with a diffeomorphism $\mathfrak{G}$. However, we prefer the formulation in (4.8) since it allows us to work on a parameter-independent domain, where discretization elements and normal vectors are fixed, and additionally casts a clear strategy for model order reduction, as we shall describe below.

4.2.2 Implementation

We rewrite the system (4.8) in terms of bilinear and linear forms. The weak formulation reads: 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

$$\begin{aligned}\mathcal{A}(\mathbf{v}_h,\boldsymbol{\kappa}) - \mathcal{B}(\boldsymbol{G}\mathbf{e}_h,\boldsymbol{\kappa}) - \mathcal{C}(\boldsymbol{G}\widehat{\mathbf{e}}_h,\boldsymbol{\kappa}) = 0,\\ \mathcal{B}(\boldsymbol{G}\mathbf{v}_h,\boldsymbol{\xi}) + \mathcal{K}(\boldsymbol{G}\mathbf{v}_h,\boldsymbol{\xi}) + \mathcal{D}(\boldsymbol{G}\mathbf{e}_h,\boldsymbol{\xi}) - \omega^2\mathcal{A}_\varepsilon(\mathbf{e}_h,\boldsymbol{\xi}) - \mathcal{E}(\boldsymbol{G}\widehat{\mathbf{e}}_h,\boldsymbol{\xi}) = 0,\\ -\mathcal{R}(\boldsymbol{G}\mathbf{v}_h,\boldsymbol{\mu}) - \mathcal{L}(\boldsymbol{G}\mathbf{e}_h,\boldsymbol{\mu}) + \mathcal{M}(\boldsymbol{G}\widehat{\mathbf{e}}_h,\boldsymbol{\mu}) = \mathcal{F}(\boldsymbol{\mu}),\end{aligned} \tag{4.9}$$

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

$$
\begin{aligned}
&\mathcal{A}(\mathbf{v}, \boldsymbol{\kappa}) = (\mathbf{v}, \boldsymbol{\kappa})_{\mathcal{T}_h}, && \mathcal{A}_\varepsilon(\mathbf{e}, \boldsymbol{\xi}) = (\varepsilon \mathbf{e}, \boldsymbol{\xi})_{\mathcal{T}_h}, \\
&\mathcal{B}(\boldsymbol{G}\mathbf{e}, \boldsymbol{\kappa}) = (\boldsymbol{G}\mathbf{e}, \nabla \times \boldsymbol{\kappa})_{\mathcal{T}_h}, && \mathcal{C}(\boldsymbol{G}\widehat{\mathbf{e}}, \boldsymbol{\kappa}) = \langle \boldsymbol{G}\widehat{\mathbf{e}}, \boldsymbol{\kappa} \times \mathbf{n} \rangle_{\partial \mathcal{T}_h}, \\
&\mathcal{K}(\boldsymbol{G}\mathbf{v}, \boldsymbol{\xi}) = \langle \boldsymbol{G}\mathbf{v}, \boldsymbol{\xi} \times \mathbf{n} \rangle_{\partial \mathcal{T}_h}, && \mathcal{D}(\boldsymbol{G}\mathbf{e}, \boldsymbol{\xi}) = \langle \tau_t \boldsymbol{G}\mathbf{e}, \mathbf{n} \times \boldsymbol{\xi} \times \mathbf{n} \rangle_{\partial \mathcal{T}_h}, \\
&\mathcal{E}(\boldsymbol{G}\widehat{\mathbf{e}}, \boldsymbol{\xi}) = \langle \tau_t \boldsymbol{G}\widehat{\mathbf{e}}, \mathbf{n_r} \times \boldsymbol{\xi} \times \mathbf{n_r} \rangle_{\partial \mathcal{T}_h}, && \mathcal{R}(\boldsymbol{G}\mathbf{v}, \boldsymbol{\mu}) = \langle \boldsymbol{G}\mathbf{v}, \boldsymbol{\mu} \times \mathbf{n_r} \rangle_{\partial \mathcal{T}_h}, \\
&\mathcal{L}(\boldsymbol{G}\mathbf{e}, \boldsymbol{\mu}) = \langle \tau_t \boldsymbol{G}\mathbf{e}, \boldsymbol{\mu} \rangle_{\partial \mathcal{T}_h}, && \mathcal{M}(\boldsymbol{G}\widehat{\mathbf{e}}, \boldsymbol{\mu}) = \langle \widetilde{\tau}_t \boldsymbol{G}\widehat{\mathbf{e}}, \boldsymbol{\mu} \rangle_{\partial \mathcal{T}_h}, \\
&\mathcal{F}(\boldsymbol{\mu}) = \mathcal{F}_D(\boldsymbol{\mu}) + \mathcal{F}_N(\boldsymbol{\mu}) + \mathcal{F}_{SM}(\boldsymbol{\mu}) &&
\end{aligned}
$$

where the boundary form arises from (4.7). For the discretization of the above system, we introduce $\varphi_i,\ 1 \leq i \leq I$ to be the basis functions of $\mathcal{P}^p(T)$ and $\phi_k,\ 1 \leq k \leq K$ the basis functions of $\mathcal{P}^p(F)$. The tangent vectors are defined as in Chapter 3, although they are obviously computed using the normal vectors from the reference discretization.

Bearing in mind that the 3x3 deformation tensor $\boldsymbol{G}$ is symmetric, we derive the contribution from an arbitrary element T to the global system. Furthermore, we shall make use of the identity $\boldsymbol{\mu} = \mathbf{n_r} \times \boldsymbol{\mu} \times \mathbf{n_r}$, that is $\boldsymbol{\mu}$ is spanned by $\mathbf{t}_1$ and $\mathbf{t}_2$. The elemental matrices for element T have the following representation:

$$
\begin{aligned}
\mathbb{A}^t &= \begin{bmatrix} \mathbb{A}_0^t & 0 & 0 \\ 0 & \mathbb{A}_0^t & 0 \\ 0 & 0 & \mathbb{A}_0^t \end{bmatrix} & \mathbb{B}^t &= \begin{bmatrix} \mathbb{B}_{11}^t & \mathbb{B}_{12}^t & \mathbb{B}_{13}^t \\ \mathbb{B}_{21}^t & \mathbb{B}_{22}^t & \mathbb{B}_{23}^t \\ \mathbb{B}_{31}^t & \mathbb{B}_{32}^t & \mathbb{B}_{33}^t \end{bmatrix} & \mathbb{K}^t &= \begin{bmatrix} \mathbb{K}_{11}^t & \mathbb{K}_{12}^t & \mathbb{K}_{13}^t \\ \mathbb{K}_{21}^t & \mathbb{K}_{22}^t & \mathbb{K}_{23}^t \\ \mathbb{K}_{31}^t & \mathbb{K}_{32}^t & \mathbb{K}_{33}^t \end{bmatrix} \\
\mathbb{C}^t &= \begin{bmatrix} \mathbb{C}_{11}^t & \mathbb{C}_{12}^t \\ \mathbb{C}_{21}^t & \mathbb{C}_{22}^t \\ \mathbb{C}_{31}^t & \mathbb{C}_{32}^t \end{bmatrix} & \mathbb{D}^t &= \begin{bmatrix} \mathbb{D}_{11}^t & \mathbb{D}_{12}^t & \mathbb{D}_{13}^t \\ \mathbb{D}_{21}^t & \mathbb{D}_{22}^t & \mathbb{D}_{23}^t \\ \mathbb{D}_{31}^t & \mathbb{D}_{32}^t & \mathbb{D}_{33}^t \end{bmatrix} & \mathbb{E}^t &= \begin{bmatrix} \mathbb{E}_{11}^t & \mathbb{E}_{12}^t \\ \mathbb{E}_{21}^t & \mathbb{E}_{22}^t \\ \mathbb{E}_{31}^t & \mathbb{E}_{32}^t \end{bmatrix} \\
\mathbb{R}^t &= \begin{bmatrix} \mathbb{R}_{11}^t & \mathbb{R}_{12}^t & \mathbb{R}_{13}^t \\ \mathbb{R}_{21}^t & \mathbb{R}_{22}^t & \mathbb{R}_{23}^t \end{bmatrix} & \mathbb{L}^t &= \begin{bmatrix} \mathbb{L}_{11}^t & \mathbb{L}_{12}^t & \mathbb{L}_{13}^t \\ \mathbb{L}_{21}^t & \mathbb{L}_{22}^t & \mathbb{L}_{23}^t \end{bmatrix} & \mathbb{M}^t &= \begin{bmatrix} \mathbb{M}_{11}^t & \mathbb{M}_{12}^t \\ \mathbb{M}_{21}^t & \mathbb{M}_{22}^t \end{bmatrix}
\end{aligned}
\tag{4.10}
$$

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 dimensionality 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

$$\begin{aligned}
\mathbb{B}^t_{1d,ij} &= (\partial_z\varphi_i, \boldsymbol{G}_{2d}\,\varphi_j)_T - (\partial_y\varphi_i, \boldsymbol{G}_{3d}\,\varphi_j)_T\ ,\\
\mathbb{B}^t_{2d,ij} &= (\partial_x\varphi_i, \boldsymbol{G}_{3d}\,\varphi_j)_T - (\partial_z\varphi_i, \boldsymbol{G}_{1d}\,\varphi_j)_T\ ,\\
\mathbb{B}^t_{3d,ij} &= (\partial_y\varphi_i, \boldsymbol{G}_{1d}\,\varphi_j)_T - (\partial_x\varphi_i, \boldsymbol{G}_{2d}\,\varphi_j)_T.
\end{aligned} \tag{4.11}$$

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

$$\kappa_{1d} = \boldsymbol{G}_{3d}n_2 - \boldsymbol{G}_{2d}n_3\ ,\quad \kappa_{2d} = \boldsymbol{G}_{1d}n_3 - \boldsymbol{G}_{3d}n_1\ ,\quad \kappa_{3d} = \boldsymbol{G}_{2d}n_1 - \boldsymbol{G}_{1d}n_2\ , \tag{4.12}$$

$$\begin{aligned}
\delta_{1d} &= \boldsymbol{G}_{1d}(1-n_1^2) - n_1n_3\boldsymbol{G}_{3d} - n_1n_2\boldsymbol{G}_{2d}\ ,\\
\delta_{2d} &= \boldsymbol{G}_{2d}(1-n_2^2) - n_1n_2\boldsymbol{G}_{1d} - n_2n_3\boldsymbol{G}_{3d}\ ,\\
\delta_{3d} &= \boldsymbol{G}_{3d}(1-n_3^2) - n_1n_3\boldsymbol{G}_{1d} - n_2n_3\boldsymbol{G}_{2d}\ .
\end{aligned} \tag{4.13}$$

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

$$\gamma_{1b} = \widehat{t}_{b3}n_2 - \widehat{t}_{b2}n_3\ ,\quad \gamma_{2b} = \widehat{t}_{b1}n_3 - \widehat{t}_{b3}n_1\ ,\quad \gamma_{3b} = \widehat{t}_{b2}n_1 - \widehat{t}_{b1}n_2\ , \tag{4.14}$$

$$\begin{aligned}
\epsilon_{1b} &= \widehat{t}_{b1}(1-n_1^2) - n_1n_3\widehat{t}_{b3} - n_1n_2\widehat{t}_{b2}\ ,\\
\epsilon_{2b} &= \widehat{t}_{b2}(1-n_2^2) - n_1n_2\widehat{t}_{b1} - n_2n_3\widehat{t}_{b3}\ ,\\
\epsilon_{3b} &= \widehat{t}_{b3}(1-n_3^2) - n_1n_3\widehat{t}_{b1} - n_2n_3\widehat{t}_{b2}\ .
\end{aligned} \tag{4.15}$$

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

The linear form can be defined as $\mathbf{F}^t = [\mathbf{F}^t_1; \mathbf{F}^t_2]$, where the components of the vector are given by

$$\mathbf{F}^t_{a,\ell} = \langle(\mathbf{t}_a\times\mathbf{n}_\mathbf{r})^*\boldsymbol{G}\mathbf{e}_D, \phi_\ell\rangle_{\partial T\cap\partial\Omega_D} - \langle\widehat{\mathbf{t}}_a^*\mathbf{v}_N, \phi_\ell\rangle_{\partial T\cap\partial\Omega_N} - \langle\mathbf{t}_b^*\mathbf{f}^{inc}, \phi_\ell\rangle_{\partial T\cap\partial\Omega_{SM}}, \tag{4.16}$$

for $a = 1, 2$.

Assembling the elemental contributions in (4.10) we arrive at the HDG system for Maxwell's

equations

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

where $\underline{\mathbf{v}}$, $\underline{\mathbf{e}}$, $\widehat{\underline{\mathbf{e}}}$ are vectors containing the values of the corresponding fields at the degrees of freedom corresponding to the discretization $\mathcal{T}_h$. Similarly as before, we locally eliminate the degrees of freedom of $\underline{\mathbf{v}}$, $\underline{\mathbf{e}}$ to obtain a system involving only the global degrees of freedom $\widehat{\underline{\mathbf{e}}}$, 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

$$\begin{bmatrix} \mathbf{v} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbb{A} & -\mathbb{B} \\ \mathbb{B}+\mathbb{K} & \mathbb{D}-\omega^2\mathbb{A}_\varepsilon \end{bmatrix}^{-1} \begin{bmatrix} \mathbb{C} \\ \mathbb{E} \end{bmatrix} \widehat{\underline{\mathbf{e}}} = \mathbb{Z}\widehat{\underline{\mathbf{e}}}. \tag{4.17}$$

Finally, the system involving only the global unknowns is given by

$$\left(\mathbb{M} + \begin{bmatrix} \mathbb{R} & \mathbb{L} \end{bmatrix} \mathbb{Z}\right) \widehat{\underline{\mathbf{e}}} = \mathbf{F} \tag{4.18}$$

This procedure characterizes the solution to (3.6) in terms of $\widehat{\mathbf{e}}_h$, which consists of only two components defined on the global faces. As a consequence, the HDG method exhibits less globally coupled degrees of freedom than other DG methods. If $\mathbb{Z}$ is stored element-wise the local variables can be recovered in parallel using (4.17) at a cost much smaller than that of solving (4.18). Finally, the physical variables may be recovered from the transformed variables through

$$\mathbf{V}_h = g^{-1}\mathcal{G}\mathbf{v}_h, \qquad \mathbf{E}_h = g^{-1}\mathcal{G}\mathbf{e}_h. \tag{4.19}$$

4.2.3 Empirical interpolation for Maxwell's equations

The formulation and implementation introduced above describe the procedure for computing solutions to Maxwell's equations for deformable domains. It is clear that the strength of the reference domain formulation arises from its combination with reduced order modeling strategies, enabling the fast evaluation of the system at multiple geometric configurations. Nevertheless, considering geometric parameters for the reduced order model gives rise to further complications.

The more obvious one is that affine parametric dependence is lost with respect to $\boldsymbol{\theta}$, because even if $\mathfrak{G}$ were affine in $\boldsymbol{\theta}$, the tensor $\boldsymbol{G}$ most surely is not. Hence, we must resort to advanced interpolation techniques to regain an affine expression for the weak formulation. Additionally, the new weak formulation derived in Section 4.1.1 is no longer valid for the reference domain Maxwell's equations. That is, eliminating the magnetic field from (4.8a) and substituting it into (4.8b) and (4.8c) incurs higher order products of $\boldsymbol{G}$, impacting computations at both offline and online stages. To circumvent this limitation, we must use the weak formulation (4.8), as well as all field variables to develop the ROM.

The first step is to interpolate the bilinear forms (4.11)-(4.16) in the weak formulation. A widespread technique in function interpolation is DEIM [37], that will be used for the examples in this dissertation. The idea behind DEIM is to approximate a nonaffine parametrized function $f(x, \boldsymbol{\theta})$ by a weighted combination of orthogonal spatial functions $\Phi = [\phi_1(x), \ldots, \phi_Q(x)]$ such that

$$f(x, \boldsymbol{\theta}) \approx f_{EI}(x, \boldsymbol{\theta}) = \Phi(x)\mathbf{c}(\boldsymbol{\theta}), \tag{4.20}$$

for a certain parameter-dependent coefficient vector $\mathbf{c}(\boldsymbol{\theta})$. In addition, the DEIM treats the spatial variable x, defined on a bounded domain, as a finite discrete set of points within the domain. The basis functions in Φ arise from applying POD to a matrix of, for instance, I snapshots evaluated on the discrete set x, that is $[f(x, \boldsymbol{\theta}_1), \ldots, f(x, \boldsymbol{\theta}_I)]$. Nonetheless, expression (4.20) represents an overdetermined system, thus further constraints are needed to compute the coefficients. Starting from the first basis function ϕ_1 of Φ, the DEIM iteratively selects Q discrete spatial point that maximize the error of the current interpolation, that is at step q we seek the spatial index that maximizes the residual $r_q = \phi_q - [\phi_1, \ldots, \phi_{q-1}]\,\mathbf{c}(\boldsymbol{\theta})$, and augment the basis. The result is a set of Q spatial indices $\mathcal{I}$ that give rise to a $Q \times Q$ system, namely

$$f_{EI}(x_{\mathcal{I}}, \boldsymbol{\theta}) = \Phi(x_{\mathcal{I}})\mathbf{c}(\boldsymbol{\theta}), \tag{4.21}$$

to solve for $\mathbf{c}(\boldsymbol{\theta})$ for a new $\boldsymbol{\theta}$. It should be remarked that, for a given nonaffine function, the matrix $\Phi(x_{\mathcal{I}})$ does not depend on $\boldsymbol{\theta}$, thus its LU factorization can be precomputed and stored to economize the evaluation of coefficients. We refer the reader to [37] for details of the algorithm.

In order to apply the DEIM to the weak formulation (4.8), we first need to identify the

nonaffine 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 nonaffine functions can be readily identified from (4.11)-(4.16), 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. In Table 4.3, we summarize the relevant information for applying the DEIM to the various bilinear forms in (4.9). Note that each nonaffine function that is interpolated may require a different amount of orthogonal basis functions to meet the DEIM accuracy requirements.

Bilinear/Linear forms	Spatial points	Nonaffine functions	# nonaffine terms
$\mathcal{B}$	Volume	$\boldsymbol{G}_{cd},\ 1 \le c \le d \le 3$	$\sum_{i=1}^{6} Q_i^B$
$\mathcal{C}$	Face	$\gamma_{cb},\ 1 \le c \le 3,\ 1 \le b \le 2$	$\sum_{i=1}^{6} Q_i^C$
$\mathcal{E}$	Face	$\epsilon_{cb},\ 1 \le c \le 3,\ 1 \le b \le 2$	$\sum_{i=1}^{6} Q_i^E$
$\mathcal{K}$	Face	$\kappa_{cd},\ 1 \le c, d \le 3$	$\sum_{i=1}^{9} Q_i^K$
$\mathcal{D}$	Face	$\delta_{cd},\ 1 \le c, d \le 3$	$\sum_{i=1}^{9} Q_i^D$
$\mathcal{R}$	Face	$[\boldsymbol{G}(\mathbf{t}_a \times \mathbf{n}_\mathbf{r})]_d,\ 1 \le a \le 2,\ 1 \le d \le 3,$	$\sum_{i=1}^{6} Q_i^R$
$\mathcal{L}$	Face	$[\boldsymbol{G}\mathbf{t}_a]_d,\ 1 \le a \le 2,\ 1 \le d \le 3$	$\sum_{i=1}^{6} Q_i^L$
$\mathcal{M}$	Face	$\mathbf{t}_a^* \boldsymbol{G} \mathbf{t}_b,\ 1 \le a, b \le 2$	$\sum_{i=1}^{3} Q_i^M$
$\mathcal{F}_D$	Dirichlet boundary	$(\mathbf{t}_a \times \mathbf{n}_\mathbf{r})^* \boldsymbol{G} \mathbf{e}_D,\ 1 \le a \le 2$	$\sum_{i=1}^{2} Q_i^{F_D}$
$\mathcal{F}_N$	Neumann boundary	$\mathbf{t}_a^* \boldsymbol{G} \mathbf{v}_N,\ 1 \le a \le 2$	$\sum_{i=1}^{2} Q_i^{F_N}$

Table 4.3: Nonaffine functions for DEIM.

Similarly as before, we set the $\mathcal{N}^\star$-dimensional space to be $\boldsymbol{W}_h^\star := \boldsymbol{W}_h \times \boldsymbol{W}_h \times \boldsymbol{M}_h(\mathbf{0})$, $\mathbf{e}_h^\star := (\mathbf{v}_h, \mathbf{e}_h, \widehat{\mathbf{e}}_h)$ and $\boldsymbol{\xi}^\star := (\boldsymbol{\kappa}, \boldsymbol{\xi}, \boldsymbol{\mu})$. The system of equations in (4.9) may be compactly rewritten, for $\mathbf{e}_h^\star \in \boldsymbol{W}_h^\star$, as

$$\mathcal{A}^\star(\mathbf{e}_h^\star, \boldsymbol{\xi}^\star; (\omega, \varepsilon, \boldsymbol{\theta})) = \mathcal{F}^\star(\boldsymbol{\xi}^\star; (\omega, \varepsilon, \boldsymbol{\theta})), \qquad \forall \boldsymbol{\xi}^\star \in \boldsymbol{W}_h^\star \,.$$

Applying the DEIM approximations outlined in Table 4.3 to the bilinear and linear forms in (4.9) recovers a bilinear form $\mathcal{A}_{EI}^\star \approx \mathcal{A}^\star$ and linear functional $\mathcal{F}_{EI}^\star \approx \mathcal{F}^\star$ that are affine in

the parameters $(\omega, \varepsilon, \boldsymbol{\theta})$, namely

$$\begin{aligned}
\mathcal{A}^{\star}_{EI}(\mathbf{u}^{\star}, \boldsymbol{\xi}^{\star}; (\omega, \varepsilon, \boldsymbol{\theta})) =& \mathcal{A}(\mathbf{p}, \boldsymbol{\kappa}) - \sum_{i=1}^{6} \sum_{q=1}^{Q_i^B} [\mathcal{B}_q(\boldsymbol{G}\mathbf{u}, \boldsymbol{\kappa}) + \mathcal{B}_q(\boldsymbol{G}\mathbf{p}, \boldsymbol{\xi})] \\
& - \omega^2 \mathcal{A}_{\varepsilon}(\mathbf{u}, \boldsymbol{\xi}) - \sum_{i=1}^{6} \sum_{q=1}^{Q_i^C} \mathcal{C}_q(\boldsymbol{G}\boldsymbol{\nu}_h, \boldsymbol{\kappa}) \\
& + \sum_{i=1}^{9} \sum_{q=1}^{Q_i^K} \mathcal{K}_q(\boldsymbol{G}\mathbf{p}, \boldsymbol{\xi}) + \tau_t \sum_{i=1}^{9} \sum_{q=1}^{Q_i^D} \mathcal{D}_q(\boldsymbol{G}\mathbf{u}, \boldsymbol{\xi}) \\
& - \tau_t \sum_{i=1}^{6} \sum_{q=1}^{Q_i^E} \mathcal{E}_q(\boldsymbol{G}\boldsymbol{\nu}_h, \boldsymbol{\xi}) - \sum_{i=1}^{6} \sum_{q=1}^{Q_i^R} \mathcal{R}_q(\boldsymbol{G}\mathbf{p}, \boldsymbol{\mu}) \\
& - \tau_t \sum_{i=1}^{6} \sum_{q=1}^{Q_i^L} \mathcal{L}_q(\boldsymbol{G}\mathbf{u}, \boldsymbol{\mu}) + \widetilde{\tau}_t \sum_{i=1}^{3} \sum_{q=1}^{Q_i^M} \mathcal{M}_q(\boldsymbol{G}\boldsymbol{\nu}, \boldsymbol{\mu}) \\
\mathcal{F}^{\star}_{EI}(\boldsymbol{\xi}^{\star}; (\omega, \varepsilon, \boldsymbol{\theta})) =& \sum_{i=1}^{2} \sum_{q=1}^{Q_i^{F_D}} \mathcal{F}_{D,q}(\boldsymbol{\mu}; \boldsymbol{\theta}) + \sum_{i=1}^{2} \sum_{q=1}^{Q_i^{F_N}} \mathcal{F}_{N,q}(\boldsymbol{\mu}; \boldsymbol{\theta}) \\
& + \mathcal{F}_{SM}(\boldsymbol{\mu}; (\omega, \varepsilon)),
\end{aligned}$$

for all $\mathbf{u}^{\star} = (\mathbf{p}, \mathbf{u}, \boldsymbol{\nu}) \in \boldsymbol{W}_h^{\star}$, and set the HDG system with DEIM approximations as

$$\mathcal{A}^{\star}_{EI}(\mathbf{e}_h^{\star}, \boldsymbol{\xi}^{\star}; (\omega, \varepsilon, \boldsymbol{\theta})) = \mathcal{F}^{\star}_{EI}(\boldsymbol{\xi}^{\star}; (\omega, \varepsilon, \boldsymbol{\theta})), \qquad \forall \boldsymbol{\xi}^{\star} \in \boldsymbol{W}_h^{\star}\,. \tag{4.22}$$

Additionally, we define an inner product for the approximation space $\boldsymbol{W}_h^{\star}$ as

$$(\mathbf{u}^{\star}, \boldsymbol{\xi}^{\star})_{\boldsymbol{W}^{\star}} = (\mathbf{p}, \boldsymbol{\kappa})_{\mathcal{T}_h} + (\mathbf{u}, \boldsymbol{\xi})_{\mathcal{T}_h} + \langle \boldsymbol{\nu}, \boldsymbol{\mu} \rangle_{\partial \mathcal{T}_h},$$

which also defines an induced norm $\|\boldsymbol{\xi}^{\star}\|_{\boldsymbol{W}^{\star}} = \sqrt{(\boldsymbol{\xi}^{\star}, \boldsymbol{\xi}^{\star})_{\boldsymbol{W}^{\star}}}$. Again, even though the formulation involves all the field variables, to compute the snapshots we leverage the structure of the HDG method and solve a linear system only for $\widehat{\mathbf{e}}_h$. We then recover the local field variables at the element level using the Schur complement decomposition.

4.2.4 POD formulation

The procedure to compute a POD basis from a matrix of snapshots is analogous to the regular Maxwell's case. The hierarchical POD space $\boldsymbol{W}_N^\star$ is similarly defined as

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

for the orthonormalized basis functions $\boldsymbol{\zeta}_n^\star \in \boldsymbol{W}_h^\star$. We then perform a Galerkin projection, for a given $(\omega, \varepsilon, \boldsymbol{\theta})$, on system (4.22) to find an approximation $\mathbf{e}_N^\star(\omega, \varepsilon, \boldsymbol{\theta}) \in \boldsymbol{W}_N^\star$ satisfying

$$\mathcal{A}_{EI}^\star(\mathbf{e}_N^\star, \boldsymbol{\xi}^\star\,; (\omega, \varepsilon, \boldsymbol{\theta})) = \mathcal{F}_{EI}^\star(\boldsymbol{\xi}^\star\,; (\omega, \varepsilon, \boldsymbol{\theta})), \qquad \forall \boldsymbol{\xi}^\star \in \boldsymbol{W}_N^\star\,.$$

Again, the solution of the $N \times N$ system above are the trial coefficients $\{\lambda_n^\star\}_{n=1}^N$, that is $\mathbf{e}_N^\star = \sum_{n=1}^N \lambda_n^\star \boldsymbol{\zeta}_n^\star$.

4.2.5 Computational strategy

The linearity and affine parametric dependence of the problem allow for a similar offline-online decomposition strategy to the regular Maxwell's case, albeit at a larger cost as a consequence of empirical interpolation. Formation and storage of the parameter-independent matrices and vectors in the offline stage is considerably more expensive, since each bilinear and linear form subject to DEIM incurs a formation and storage cost Q times larger than before, where Q is the total number of nonaffine terms for that particular form, given in the last column of Table 4.3. Finally, in the offline stage, we also store the DEIM indices $\mathcal{I}$, the resulting square matrices $\Phi(x_{\mathcal{I}})$ and its LU factorizations, see (4.21), so as to expedite the online recovery of the parameter-dependent coefficients $\mathbf{c}(\boldsymbol{\theta})$.

The online stage evaluates, for a new tuple $(\omega, \varepsilon, \boldsymbol{\theta})$, the nonaffine functions at the DEIM indices with a cost $\mathcal{O}(\mathcal{Q}\mathcal{N}_{\mathcal{I}}^\star)$ and the DEIM coefficients $\mathbf{c}(\boldsymbol{\theta})$ in $\mathcal{O}(2\mathcal{Q}^{\star 2})$ operations, where $\mathcal{Q}$ and $\mathcal{Q}^{\star 2}$ are given by

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

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

Evaluating the DEIM functions has $\mathcal{N}^\star-$dependence. Nevertheless, the dimension of the set of DEIM indices $\mathcal{N}_\mathcal{J}^\star$ where the functions are evaluated is usually much smaller than the discrete set of Gaussian quadrature points $\mathbf{x}_\xi$, thus $\mathcal{N}_\mathcal{J}^\star \ll \mathcal{N}^\star$. Computing the trial coefficients $\{\lambda_n^\star(\omega,\varepsilon)\}_{n=1}^N$ requires complexity $\mathcal{O}\left(N^3+\mathcal{Q}N^2\right)$, independent of the dimension $\mathcal{N}^\star$ of the HDG approximation space.

In a similar fashion, the dependence on $\mathcal{N}^\star$ appears when computing $\mathbf{v}_N$, $\mathbf{e}_N$, for a complexity of $\mathcal{O}\left(4\mathcal{N}^\star N\right)$. The actual approximate electromagnetic fields $\mathbf{V}_N$, $\mathbf{E}_N$ are obtained applying (4.19) to $\mathbf{v}_N$, $\mathbf{e}_N$ elementwise. However, the idea of localized integrals to evaluate QoI introduced before is still valid, thus we expect a cost much lower than the theoretical complexity.

4.2.6 Numerical results

In this section, we extend the previous periodic annular example, with geometry defined in Fig. 3-4, to demonstrate the effectivity of ROM for geometry deformations. The objective is to show that, when the domain is subject to deformations, instead of solving Maxwell's equations on the deformed discretization it is desirable to solve the modified Maxwell's equations (4.6) on the original discretization.

We analyze the effect of modifying the radius and gapsize of the rings at maximum field enhancement, with nominal values $R = 16$ μm and $w = 10$ nm. The modifications considered in this thesis, for both radius and gap, are homogeneous with respect to the angle of the ring. The material parameters and gold optical constants are set to $\varepsilon_{\mathrm{Al_2O_3}} = 5.5$, $n_{\mathrm{sapphire}} = 3.07$, $\hbar\omega_p = 9.02$ eV, $\hbar\gamma = 0.02678$ eV, $\varepsilon_\infty = 1$, and we study two different cases: a $\pm 30\%$ variation for the radius, that is $R \in [11.2,\ 20,8]$ μm, denoted as (R); and a $\pm 20\%$ variation for both the radius and the gap, namely $R \in [12.8,\ 19.2]$ μm and $w \in [8,\ 12]$ nm, denoted as (RG). To maximize the predictive power of the model, frequencies ranging from 0.3 to 0.9 THz are considered. Firstly, we introduce the diffeomorphism $\mathfrak{G}$ that models the deformations. The choice of $\mathfrak{G}$ is nontrivial, since we require continuity of its derivatives for the entire domain, in order to ensure the face integrals in the weak formulation are well defined. The construction that will be used is detailed in Appendix B.

The POD basis for the radius (resp. radius-gap) model is formed by computing 350 (resp.

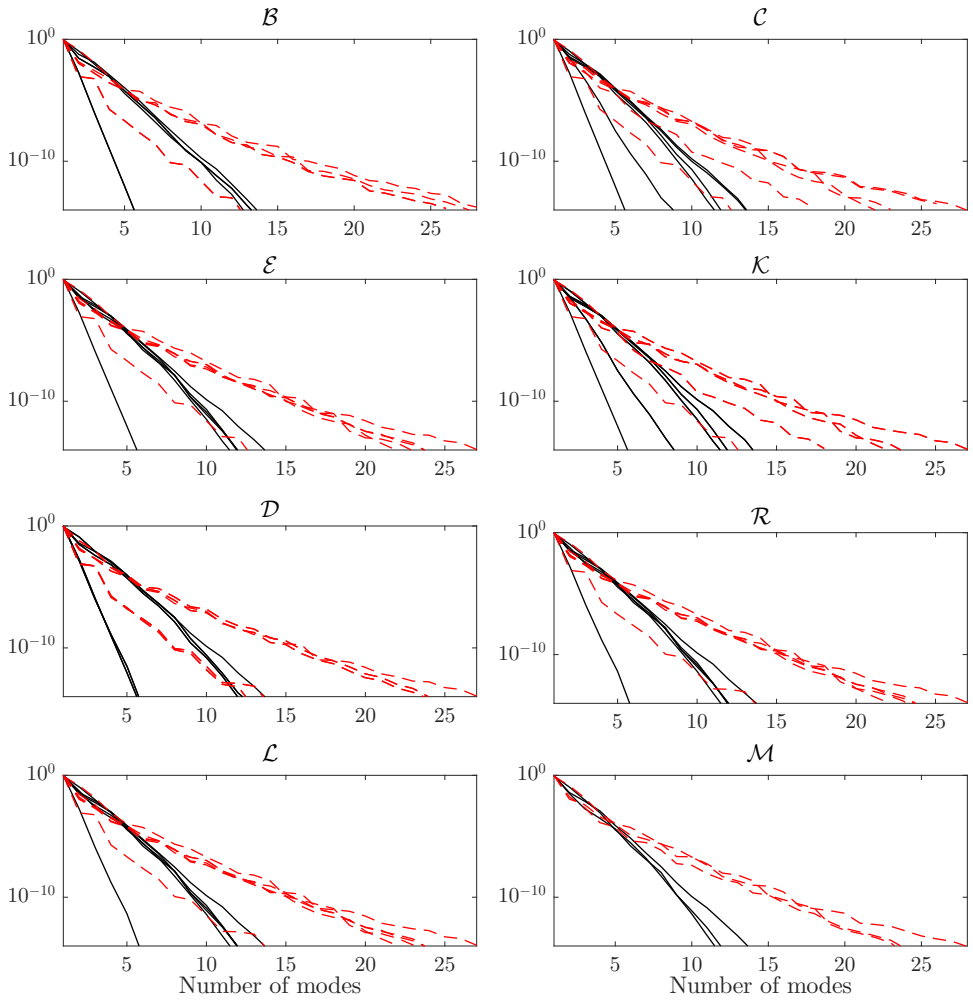


Figure 4-2: Normalized eigenvalue decay of POD applied to nonaffine functions in Table 4.3, for model (R) in solid black and (RG) in dashed red.

Model	DEIM coeff. (ms)	Assembly (ms)	Linear system (ms)	Local variables (ms)
$N = 25$ (R)	19.90	2.31	0.19	75.30
$N = 50$ (R)	20.01	7.77	0.36	75.77
$N = 75$ (R)	19.92	17.87	0.68	79.39
$N = 92$ (R)	19.89	30.07	0.90	79.03
$N = 25$ (RG)	21.93	3.52	0.20	84.21
$N = 50$ (RG)	21.05	8.54	0.36	81.85
$N = 75$ (RG)	21.77	21.92	0.69	82.48
$N = 86$ (RG)	21.91	27.29	0.83	83.17
HDG	–	9.8e4	3.00e5	9.5e2

Table 4.4: Computational wall time in milliseconds of HDG/HDG-POD for Maxwell's equations in reference domain.

600) solutions of (4.8) at different $(\omega, \boldsymbol{\theta})$ values, and then compressed on a basis of 92 (resp. 86) modes. The empirical interpolation can be performed very efficiently, since it

only requires evaluating $\mathfrak{G}(\mathbf{x}_\xi, \boldsymbol{\theta})$, its derivatives, and the subsequent nonaffine functions defined in Table 4.3. We then apply DEIM separately for each function. The eigenvalue decay associated with the POD of each nonaffine function is shown in Fig. 4-2 and we see that no more than 15 (resp. 30) modes are needed for any of the nonaffine functions.

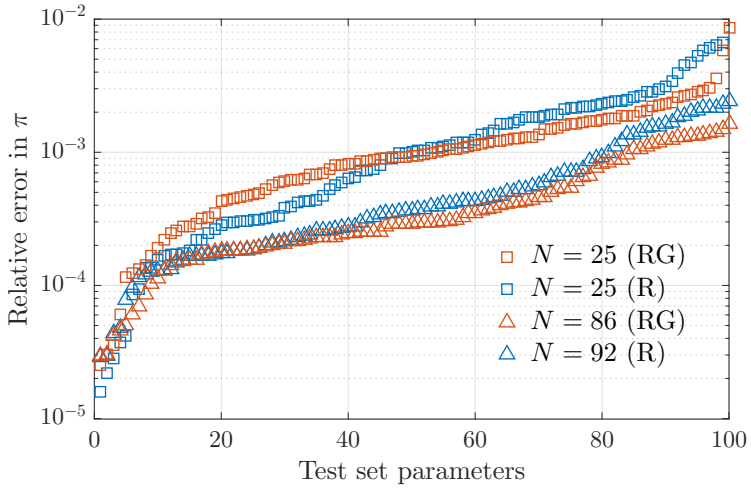
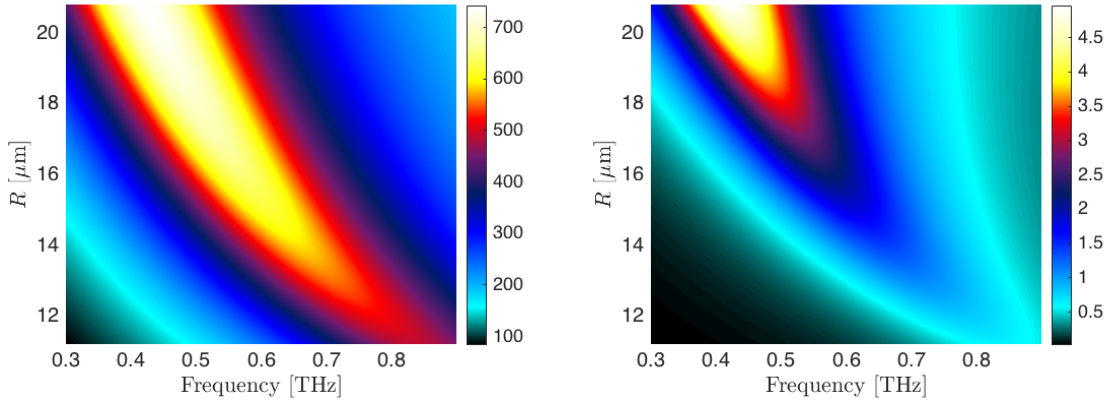


Figure 4-3: Relative error in field enhancement for test set parameters, evaluating both (R) and (RG) models at two POD sizes.



(a) Field enhancement

(b) Transmitted power (%)

Figure 4-4: Frequency and radius sweep computed with the (R) reduced order model.

Once the ROM is constructed, it can be queried in real time 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 $\mathcal{Q}$ empirical interpolation coefficients $\mathbf{c}(\boldsymbol{\theta})$ is independent of the dimension of the ROM. Its cost is mainly devoted to evaluating the nonaffine functions $f_{EI}(x_\mathcal{I}, \boldsymbol{\theta})$ in Table 4.3 at the selected indices $\mathcal{I}$. As in the time-harmonic Maxwell case, the bulk of the cost is incurred calculating the approximate electromagnetic fields from the

trial coefficients, since it involves operating with the high-fidelity dimension $\mathcal{N}^\star$. Moreover, the reference domain formulation requires non negligible additional computations, since the actual field variables need to be recovered through (4.19). Again, the benefits of employing a ROM to evaluate QoI 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 of empirical interpolation, see Table 4.4. As far as accuracy is concerned, we evaluate the field enhancement for a test set of parameters using the high-fidelity HDG model, and report the relative error for both models in Fig. 4-3. Selecting the maximum number of modes, we achieve a relative error on the order of 0.1%, although the increase in accuracy is modest as the ROM is enriched.

Finally, we go one step further and show how ROM can be leveraged to achieve a deeper understanding of this structure. For the (R) model, the field enhancement profile is shown in Fig. 4-4a as a function of both the frequency and the radius of the dielectric aperture. It is evident that reducing the radius has a significant impact in both the maximum field enhancement and the resonant frequency, both reducing the field localization and blueshifting the resonance. Analogously, the transmission power profile is collected in Fig. 4-4b, where a similar trend is identified.

The (RG) model 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. In Figs. 4-5a and 4-5c. We show, as a function of the ring radius, the value of the resonant wavelength λ^* and the maximum field enhancement π^* for the nominal gap width 10 nm (solid black). In addition, we provide intervals that correspond to the sensitivities of the resonant quantities with respect to $\pm 5\%$, $\pm 10\%$ and $\pm 20\%$ relative variation of the gap δw. Sensitivities as a function of the gap with respect to radius variations δR are obtained analogously, and are presented in Figs. 4-5e and 4-5g. 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 to relative shifts in resonant wavelength $\delta\lambda^*$ and relative variations of maximum field enhancement $\delta\pi^*$. Results are collected in Figs. 4-5b, 4-5d, 4-5f and 4-5h, such that each row in Fig. 4-5 corresponds to

one particular scenario. These mappings are obtained effortlessly once the ROM is available, and yet contribute to acquiring a more profound understanding of the device's behavior. For instance, one can endow numerical simulations with the ROM methodology to predict how manufacturing tolerances may affect the performance of the plasmonic structure, and thus facilitate the design of robust annular gaps.

4.3 Concluding remarks

In this chapter, we have extended the HDG methods introduced in the previous chapter to accommodate model order reduction using proper orthogonal decomposition techniques. In general, the material properties of the dielectrics and metals are not known precisely, since they are determined experimentally and are therefore subject to significant uncertainty. Furthermore, the extreme scales of the plasmonic devices hamper its manufacturing, hence faithful simulations require accounting for a certain degree of geometric variability. In order to circumvent this limitation, an alternative formulation of the HDG method for Maxwell's equations has been developed for a reference (parameter-independent) domain. The reference formulation, combined with discrete empirical interpolation techniques, enables us to incorporate the variables that parametrize the deformation mapping into the reduced order model.

To demonstrate the effectiveness of the approach we have constructed several ROM for the 3d periodic annular structure. The offline-online decomposition pursued leads to an accurate ROM that produces approximate field solutions and QoI in real time. This constitutes the major advantage of the HDG-POD technique, since it encapsulates multiple geometry and material parameter configurations on a single model that is extremely efficient to evaluate. At the expense of a computationally demanding offline stage, in which multiple solutions of the high-fidelity model are computed, we obtain a surrogate whereby design exploration and optimization of plasmonic devices is much more affordable.

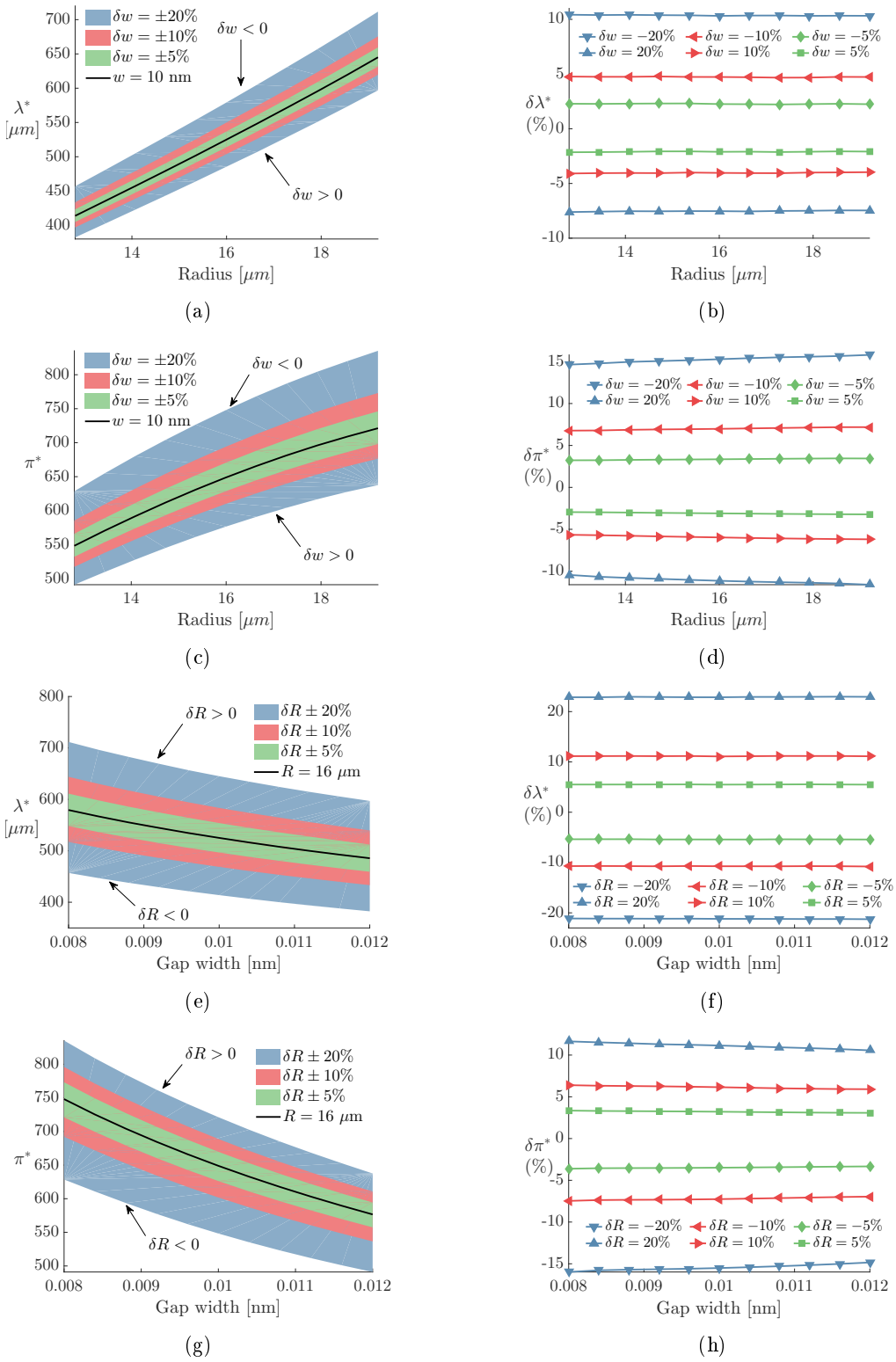


Figure 4-5: Absolute and relative sensitivities of resonant quantities on geometry modifications with (RG) model. (a)-(b) Gap variations and resonant wavelength. (c)-(d) Gap variations and maximum enhancement. (d)-(e) Radius variations and resonant wavelength. (f)-(g) Radius variations and maximum enhancement.

Chapter 5

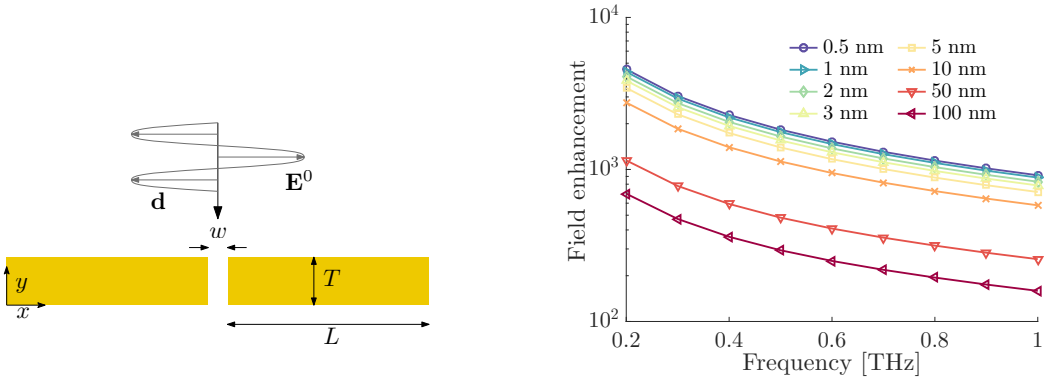
Applications

In the preceding chapters, we have introduced the HDG method as a technique to numerically simulate the interaction of EM waves with metallic nanostructures. We have also constructed a reduced order model to approximate the high-fidelity results at a fraction of the computational cost of a single, large scale simulation. In this chapter, we seek to combine the above methodologies to address several plasmonic applications. Firstly, we study the phenomenon of field enhancement saturation in a 2d slit and assess the influence of nonlocality. Secondly, we provide a parametric analysis of the annular resonator structure for the far and mid IR regimes. Thirdly, we compare numerical simulation results with experimental data reported in literature for annular structures in order to assess the performance of our methodologies and address the discrepancies that arise. Finally, we pursue the design of a concentric ring structure with enhanced transmission capabilities.

5.1 Saturation of field enhancement in nanoslit

For the first application, we study the field enhancement and transmission of a 2d nanoslit when illuminated by millimeter-long EM waves as a function of the thickness T of the metal film and the aperture w of slit. This application is particularly interesting because, at a fixed frequency, field enhancement saturates for shrinking gapsize, even without considering nonlocality or quantum effects. This phenomenon stems from pure geometrical considerations, whereby the thickness of the metal film is inversely proportional to the maximum

enhancement that can be attained. The theory is developed in [140] for a single slit modeled as a PEC, and some calculations with LRA are provided for films that are several microns thick. The study is extended here for an infinite slit patterned in a thin film, with apertures in the nanometric and subnanometric regime, which were not addressed in the original work. Furthermore, the influence of nonlocality on both the field enhancement and transmission is evaluated.

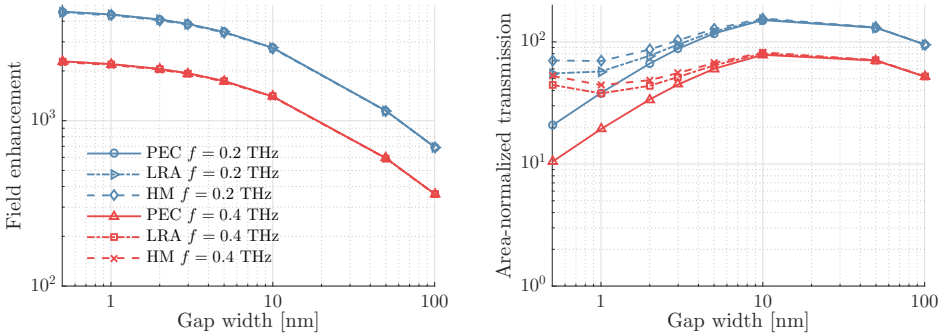


(a) (b)

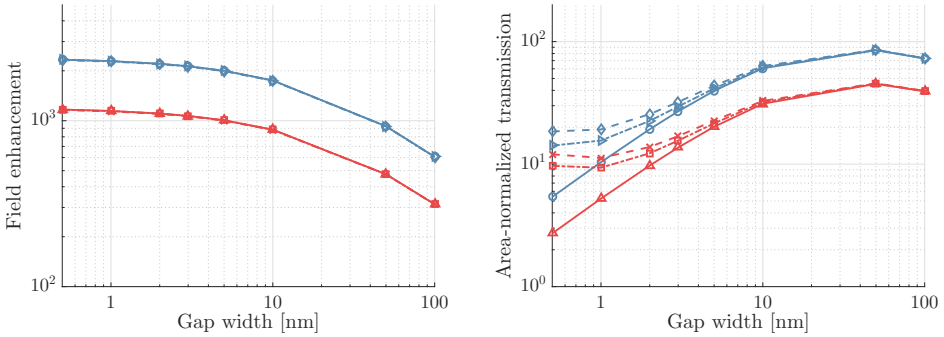
Figure 5-1: (a) Schematic of gold nanoslit with dimensions. (b) Frequency-field enhancement curves with LRA for several gap sizes and $T = 100$ nm.

The structure of interest consists of a gold film of thickness T suspended in free space with a single aperture, illuminated from above as shown in Fig. 5-1a. Radiation conditions are prescribed on the upper and lower boundaries, whereas periodic conditions are used for the lateral boundaries. We choose a lateral distance L equal to 5 mm to ensure there is no influence of the adjancent slits for the frequency regime of interest. Indeed, for the lowest frequency considered (0.2 THz) the distance between slits is larger than six wavelengths, and we numerically verify that the behavior of the structure is that of a single slit. We consider thicknesses of 100, 200 and 300 nm, and aperture widths of 0.5, 1, 2, 3, 5, 10, 50 and 100 nm, which span more than two orders of magnitude. The discretization employed is grid-converged for the smallest gap size. In Fig. 5-1b we show the field enhancement profile for several frequencies and slit apertures, computed with LRA for a 100 nm gold film, where it can be appreciated that field enhancement saturates as the gap is decreased, for a fixed frequency.

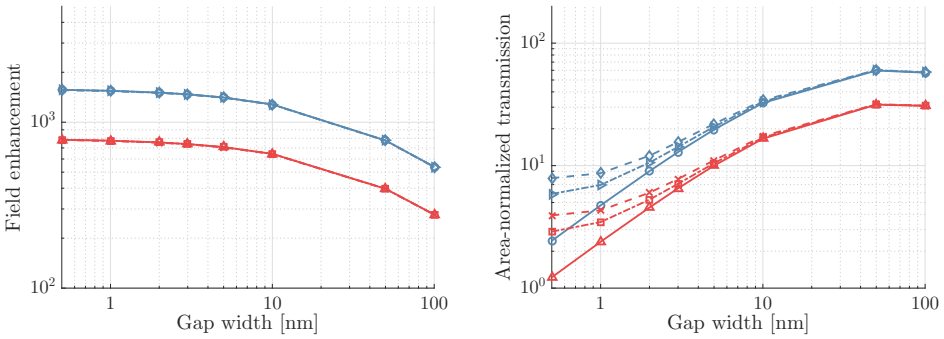
In Fig. 5-2 we show, as a function of the gap width, both the field enhancement and the optical transmission normalized by the open area fraction $w/(2L + w)$, where the metal



(a) $T = 100$ nm



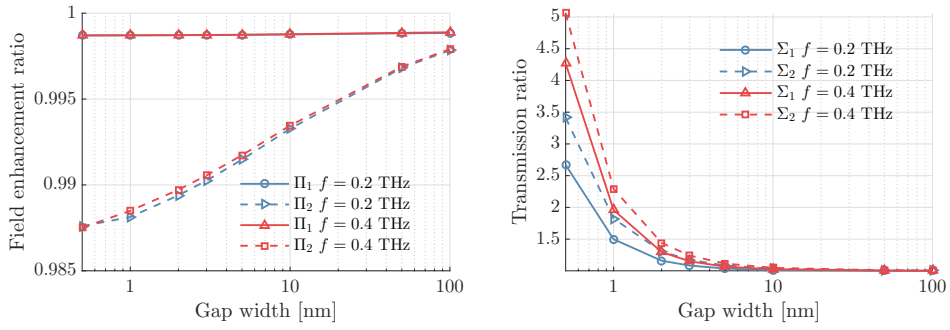
(b) $T = 200$ nm



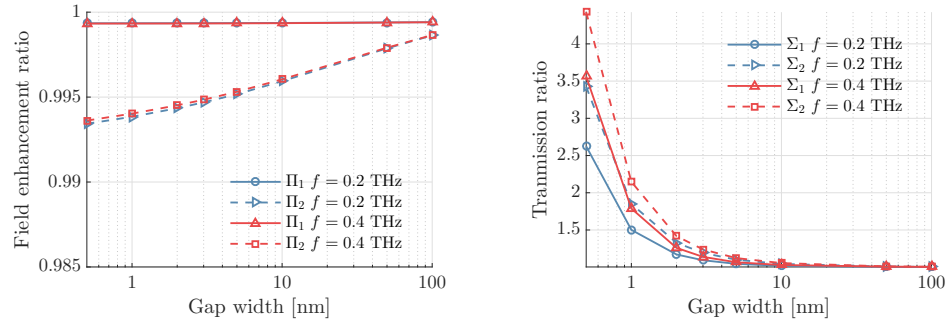
(c) $T = 300$ nm

Figure 5-2: Field enhancement and area-normalized transmission for different electron models, frequencies and metal thicknesses as a function of gap width. Legend is the same for all subfigures.

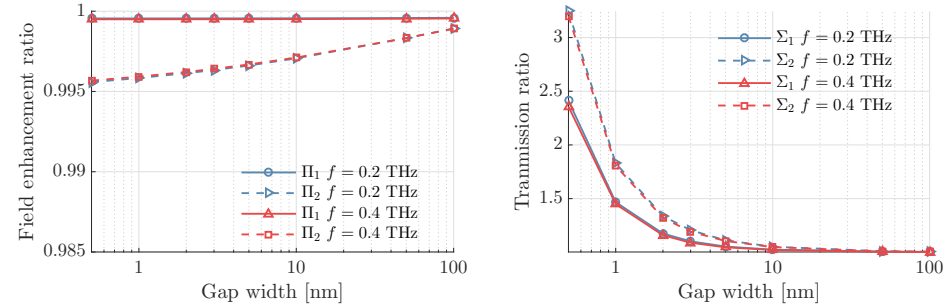
is modeled with PEC, LRA and HM at 0.2 and 0.4 THz. These results corroborate the saturation of field enhancement, observed also for the hydrodynamic model, and its inverse dependence on both the frequency and the film thickness. Conversely, the model used for light-metal interaction does have a noticeable impact on the area-normalized transmission, exhibiting increasingly large values as either the frequency is reduced or a more accurate model is used, for a fixed gap. Indeed, transmission is influenced by the effective aperture size seen by the incident field, and we have extensively shown that nonlocality broadens the



(a) $T = 100$ nm



(b) $T = 200$ nm



(c) $T = 300$ nm

Figure 5-3: Field enhancement and area-normalized transmission ratios comparing PEC with LRA and HM, for different frequencies and metal thicknesses as a function of gap width.

gap by diffusing the metal-dielectric interface. The ohmic losses within the metal explain the differences in transmission.

Finally, to better contrast the effects of the models we compute several ratios of field enhancements and transmission, namely $\Pi_1 = \pi_{LRA}/\pi_{PEC}$, $\Pi_2 = \pi_{HM}/\pi_{PEC}$ and $\Sigma_1 = \varsigma_{LRA}/\varsigma_{PEC}$, $\Sigma_2 = \varsigma_{HM}/\varsigma_{PEC}$, depicted in Fig. 5-3 as a function of the gap width. The field enhancement predicted by both PEC and LRA is essentially the same, and their ratio does not depend on the gap width. Alternatively, the discrepancy in field enhancement between LRA and HM grows for decreasing gap size and film thickness as one would expect. As far as transmission is concerned, both ratios increase as we consider smaller gaps. Indeed,

introducing losses in the metal –PEC models the metal as a lossless mirror– translates into substantial changes in the EM response of the structure at nanometric scales.

5.2 Parametric study of annular apertures

In this section, we present a parametric study of the periodic annular structure, and analyze the effects of including nonlocality for different gap sizes and frequency regimes. The geometry under study consists of an array of annular gaps with diameter D and periodicity $2D$, patterned in a 150 nm gold film, shown in Fig. 5-4.

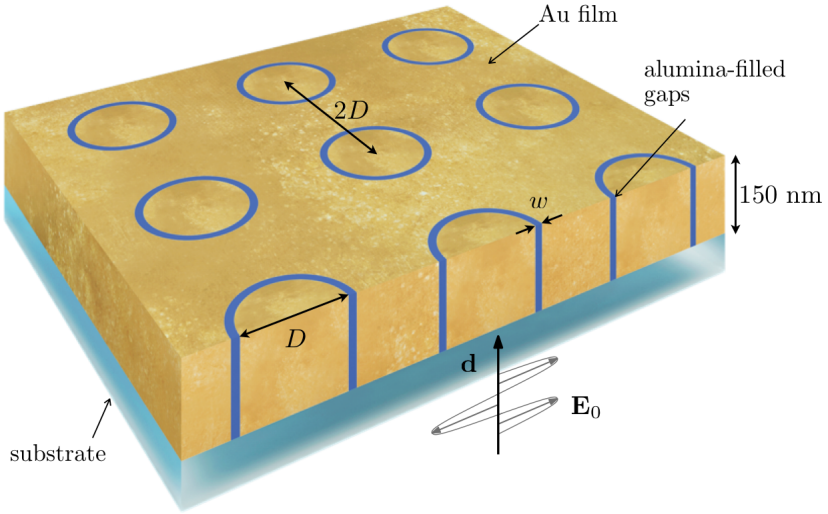


Figure 5-4: Schematic diagram of thin gold film on substrate patterned with periodic square array of alumina gaps under plane wave illumination.

5.2.1 Far infrared

We first focus on the low THz regime (0.2 to 2 THz), which requires diameters on the order of tenths of microns to excite resonances. We study four gapsizes, namely $w = 1$, 2, 5 and 10 nm, for which the response given by the different light-metal interaction models may differ significantly. We shall also investigate the influence of small variations in the radius across gapsizes and array periodicity. This is a good test to excercise the reduced order modeling capabilities proposed in this dissertation. Analogous studies can be carried out, for instance, for small variations on the gapsize or the optical constants.

The substrate for the gold film is sapphire, which is transparent at FIR, with a constant

refractive index of $n_{\text{sapphire}} = 3.07$ given by Grischkowsky *et al.* [91]. For this low frequencies, the optical constants of gold are taken from Ordal *et al.* [192, 193], with nominal values $\hbar\overline{\omega}_p = 9.02$ eV, $\hbar\overline{\gamma}_p = 0.02678$ eV and $\varepsilon_\infty = 1$. The dielectric material used for the gap is again aluminum oxide Al_2O_3 [92], and the nominal values used are $\varepsilon_{Al_2O_3} = \{3, 3, 4.4, 5.5\}$ for $w = 1, 2, 5$ and 10 nm respectively. We use the Drude permittivity to model metal response.

We develop a collection of ROM for the gapsizes above and for diameters corresponding to $D = \{12, 15, 18, 21, 24\}$ μm. In addition, for each ROM we consider a 10% variability in the radius of the ring. The ROMs are constructed using the procedures introduced in Chapter 4. That is, a predetermined number of snapshots is computed and then compressed with POD, and the resulting orthonormal basis is then used for the Galerkin projection. In Figs. 5-5a to 5-5d we show the wavelength-field enhancement profile for the various gapsizes using the corresponding nominal value of the radius, computed evaluating the ROMs.

First, we analyze the impact of the ring diameter and array periodicity on the resonant wavelength λ^* and the maximum field enhancement π^* observed. The variability of the radius is encoded in the ROM, enabling us to obtain $D - \lambda^*$ and $D - \pi^*$ curves with intervals corresponding to a $\pm 10\%$ radius variation. The results are reported in Figs. 5-5e and 5-5f, where a linear dependence between the diameter and both λ^*, π^* may be identified, which also holds for perturbations of the radius. This relationship bears similarity to the one experimentally observed for two metallic nanoparticles in [273], where the localized field enhancement in the interparticle region is proportional to the ratio between the nanoparticle's diameter and the separation, although it has never been reported for annular structures. The linear dependence is verified performing a least squares fit of the simulation data, and we report the norm of the relative error as

$$\epsilon = \left(\sum_{n=1}^{N} \left(\widetilde{y}_n - y_n \right)^2 / y_n^2 \right)^{1/2}$$

in Table 5.1, where y_n is the simulation data and $\widetilde{y}_n$ is the prediction by the fitted linear function for $N = 6$ simulation data points.

Secondly, we investigate the impact of the radius relative variation on the relative variation of both λ^* and π^* for different diameters. More specifically, for the prescribed $\delta R = \pm 10\%$

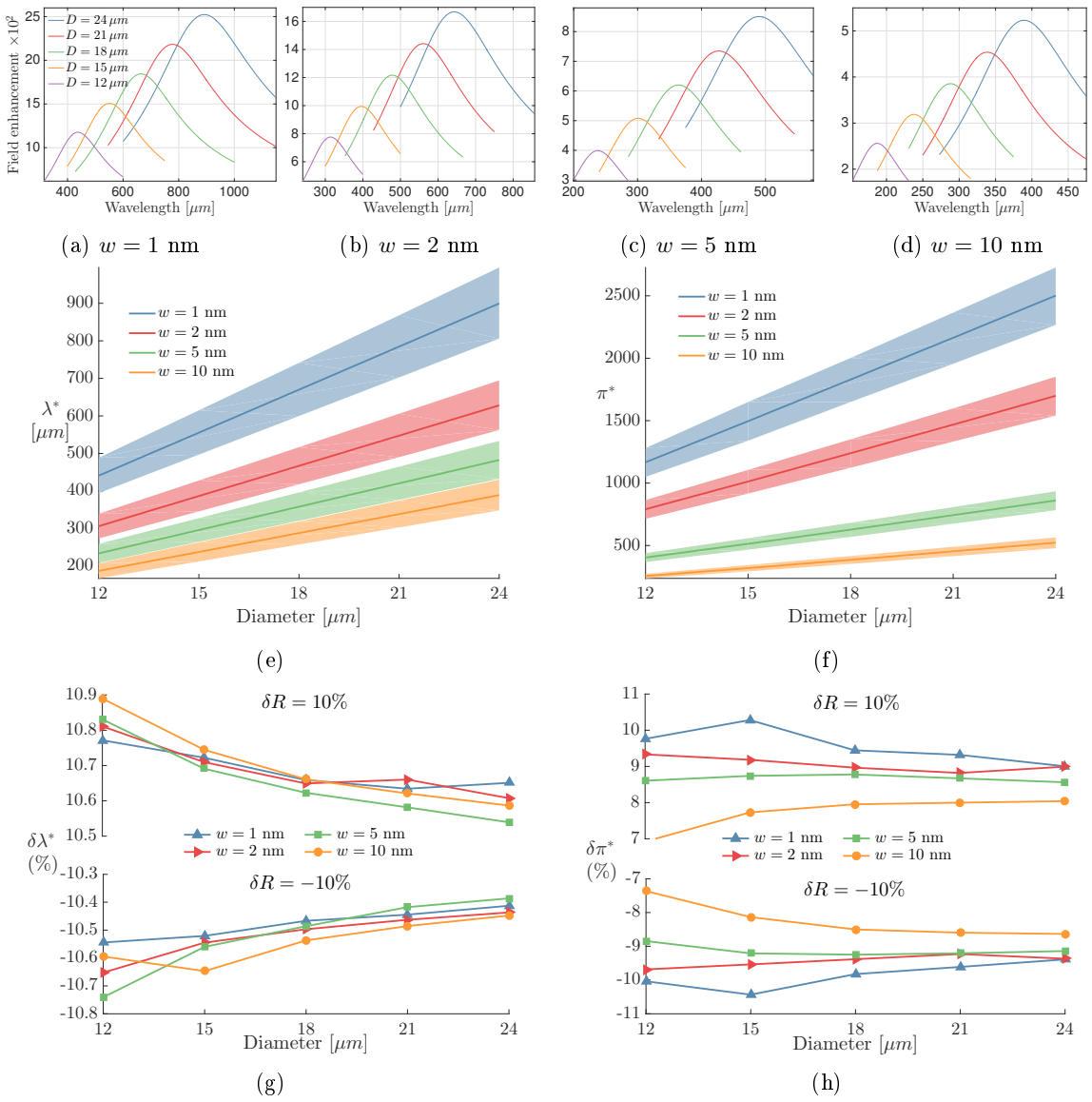


Figure 5-5: Parametric study with ROM for annular structure at FIR. (a)-(d) Wavelength - field enhancement curves for LRA computed with ROM for gapsize 1, 2, 5 and 10 nm. (e)-(f) Resonant wavelength/maximum field enhancement - diameter curves, with interval for ±10% radius variation. (g)-(h) Input-output map of relative variations as a function of diameter.

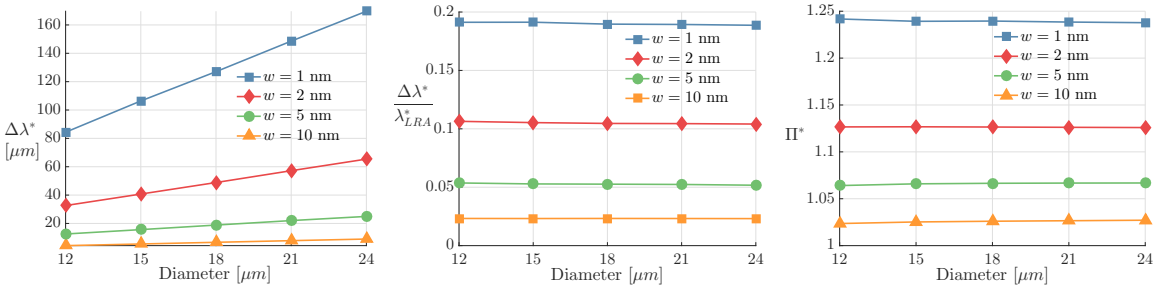
Variables \ Gap	1 nm	2 nm	5 nm	10 nm
$\epsilon(\lambda^*)$	3.7e−4	2.7e−4	9.7e−5	4.5e−4
$\epsilon(\pi^*)$	4.3e−3	5.2e−3	8.8e−3	1.2e−2

Table 5.1: ϵ of fitting for resonant wavelength and maximum field enhancement at multiple gaps.

variation, we calculate the relative variations $\delta\lambda^*$ and $\delta\pi^*$, and express this input-output map of variations as $\delta R \mapsto (\delta\lambda^*, \delta\pi^*)$. According to Figs. 5-5g and 5-5h, we observe that modifying the radius has a fairly constant impact on λ^* throughout gapsizes, slightly greater

than the relative radius input variation of 10%. The effect of radius variation on $\delta\pi^*$ is larger for decreasing gap, and is smaller than the input $\pm 10\%$ radius variation.

This example illustrates the advantages of ROM for parametric studies, since relevant and often unintuitive information can be efficiently extracted, following a computationally intensive offline stage. Data from simulations may be used to gain a deeper understanding of the phenomena of interest, as well as offering valuable insight for the design process.



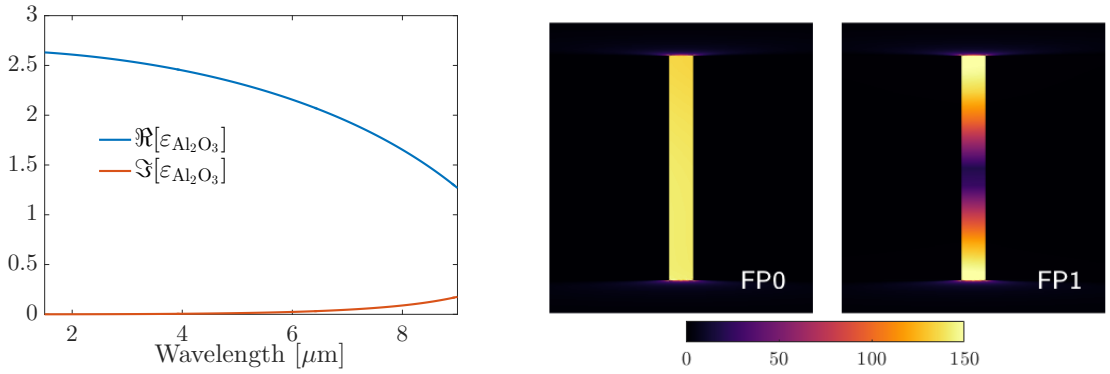
(a) Blueshift (b) Relative blueshift (c) Decay in maximum field enhancement

Figure 5-6: Effects of nonlocality for 1, 2, 5 and 10 nm gaps at far IR.

Finally, we assess the influence of nonlocality on the structure. To that end, we perform a frequency sweep solving (3.8) for the various diameters and gapsizes, without any geometric variation, and analyze the effect caused by incorporating the hydrodynamic current as the electron model on the blueshift $\Delta\lambda^* = \lambda^*_{LRA} - \lambda^*_{HM}$ and the decay in maximum field enhancement $\Pi^* = \pi^*_{LRA}/\pi^*_{HM}$. The shift in resonant wavelength, shown in Figs. 5-6a and 5-6b, is more acute for smaller gaps, and for this structure it ranges from 19% for the 1 nm gap down to 2% for the 10 nm gap. Analogously, the maximum field enhancement attained by the LRA is nearly 25% larger than that of the HM for the 1 nm gap, whereas for the 10 nm is merely 2% larger, see Fig. 5-6c. As anticipated by the theoretical results, nonlocality becomes essential to accurately predict the response of devices in the nanometric regime. In addition, we observe that both the relative wavelength blueshift and the decay in maximum field enhancement are almost constant for all diameter sizes, hence nonlocal effects are only relevant for changes in the aperture distance.

5.2.2 Mid infrared

This case corresponds to the mid infrared regime, with wavelengths from 1.5 μm to 10 μm. At these wavelengths, annular structures can sustain collective plasmon excitations for diameters in the order of hundreds of nanometers. The study is performed for 1 nm and 2 nm gap widths, and we investigate the effect of geometry variations on the extraordinary optical transmission properties of the cavity. The enhanced tranmission is caused by Fabry-Pérot (FP) resonances, which propagate along the length of the gap. Here we present results for the zeroth and first order FP modes. From the simulation perspective, the reduction in length scale mismatch renders better conditioned linear systems, thus reducing computational power requirements.



(a) (b)

Figure 5-7: (a) Real and imaginary part of Al_2O_3 permittivity at MIR, extracted from [128]. (b) Detail of $|\mathbf{E}_x|$ for 1 nm gap and 240 nm at aperture, for wavelengths $\lambda = 6.86$ μm (FP0) and $\lambda = 3.35$ μm (FP1).

For this frequency regime, we consider a substrate of silicon Si, which unlike sapphire is transparent at MIR [36] with a refractive index given by

$$n_{\text{Si}}^2 = 11.67316 + \frac{1}{\lambda^2} + \frac{0.004482633}{\lambda^2 - 1.108205^2}$$

for λ in microns. The optical constants of gold are $\hbar\overline{\omega}_p = 8.45$ eV, $\hbar\overline{\gamma}_p = 0.047$ eV and $\varepsilon_\infty = 1$, according to the experimental fitting in [190]. The dielectric material used for the gap is alumina Al_2O_3, with a refractive index obtained by interpolating the experimental data published in [128], and reproduced in Fig. 5-7a.

As before, a set of ROMs is constructed for Maxwell's equations with the Drude model, for

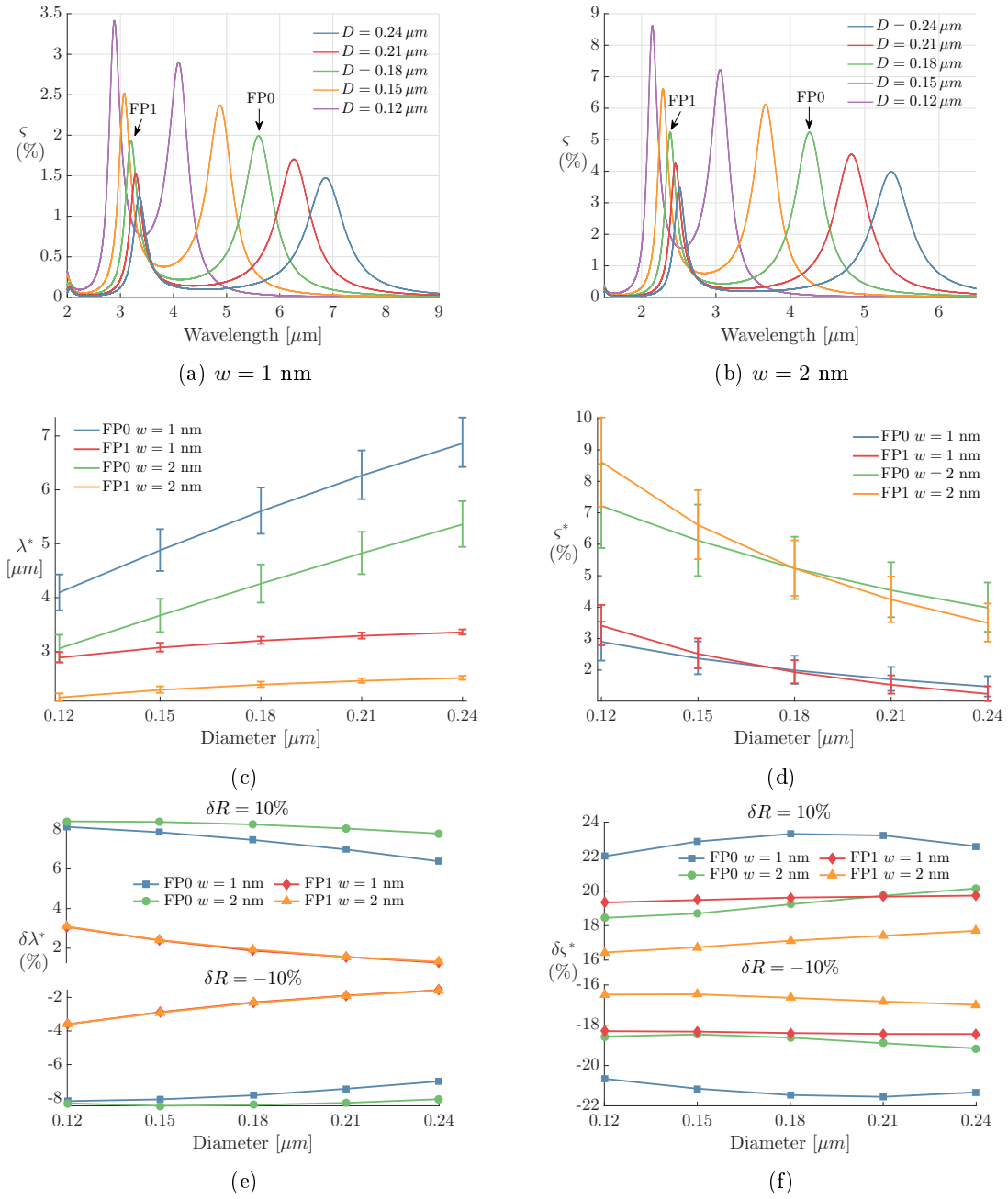


Figure 5-8: Parametric study with ROM for annular structure at MIR. (a)-(b) Wavelength - transmitted power curves for LRA computed with ROM for gapsize 1 and 2 nm. (c)-(d) Resonant wavelength/maximum transmitted power - diameter curves, with interval for ±10% radius variation. (e)-(f) Input-output map of relative variations as a function of diameter.

both gap widths and diameters $D = \{120, 150, 180, 210, 240\}$ nm, which include a ±10% variation in the radius. The transmitted power-wavelength curves are depicted in Figs. 5-8a and 5-8b, where the FP0 (longer wavelength) and FP1 (shorter wavelength) modes can be clearly identified. Both FP modes for 1 nm gap and diameter of 240 nm are reproduced in

Fig. 5-7b, where the field enhancement along the aperture is shown at their corresponding resonant frequency.

The parametric study for the annular ring at MIR is noticeably different from its FIR counterpart, and we employ an exponential function to properly fit the simulation data in Figs. 5-8c and 5-8d given by

$$\lambda^*(D) = \alpha_0 \exp\left(\alpha_1 D + \alpha_2 D^2\right).$$

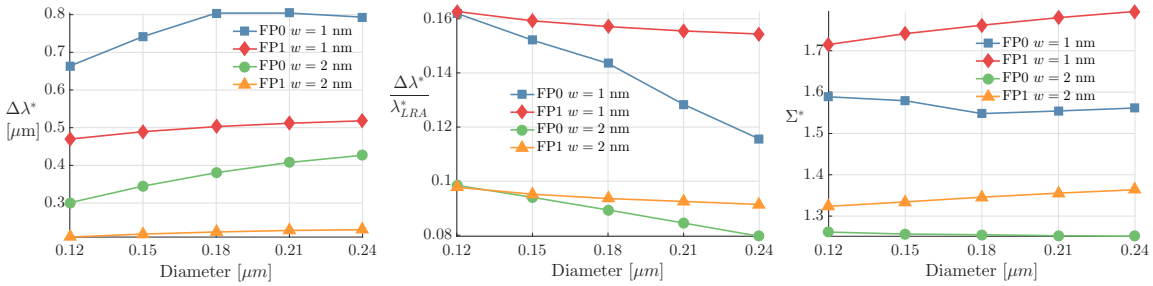
The triplet $(\alpha_0, \alpha_1, \alpha_2)$ are fitting parameters whose values are computed through least squares minimization. The norm of the relative error for the data fitting is reported in Table 5.2. As far as variations are concerned, the impact of a 10% radius variation significantly changes among modes and diameter sizes, ranging from more than 8% shift for FP0 and 120 nm diameter to less than 2% shift for FP1 and 240 nm diameter, see Fig. 5-8e. It can thus be concluded that a radius modification has a much stronger impact on the FP0 resonant wavelength than that of FP1, and that it diminishes with increasing diameter.

Variables \ Resonance	FP0 1 nm	FP1 1 nm	FP1 2 nm	FP1 2 nm
$\epsilon(\lambda^*)$	3.8e−3	2.2e−3	1.0e−3	5.7e−4
$\epsilon(\varsigma^*)$	5.9e−3	6.1e−3	1.2e−3	4.0e−3

Table 5.2: ϵ of fitting for resonant wavelength and maximum transmitted power at multiple gaps and FP resonances.

Regarding the maximum transmitted power ς^*, distinct trends can again be identified for both modes. The decay of ς^* as the diameter increases is sharper for FP1, and it is observed at a much narrower wavelength window (roughly half a micron, as opposed to around 3 microns for FP0). The input-output map of variations $\delta R \mapsto \delta\varsigma^*$ exhibits a greater impact on the FP0 modes than on FP1 modes, being rather constant for both across diameter sizes. Moreover, the variation on transmitted power is considerably amplified with respect to the variation in radius. For instance, the maximum transmission of the FP0 1 nm gap structure is subject to an alteration that exceeds ±20% when the radius changes by ±10%. This observation motivates the need to incorporate geometric modifications in the simulation and design of plasmonic devices, since manufacturing errors –which are more likely to occur for the smaller nanostructures– may lead to significant deviations in the expected perfomance

of the devices.



(a) Blueshift (b) Relative blueshift (c) Increase in transmission

Figure 5-9: Effects of nonlocality for 1, 2 nm gaps and FP0, FP1 resonances at mid IR.

To conclude the parametric analysis for the mid infrared annular nanogap, we perform hydrodynamic calculations for the diameters, gapsizes and FP resonances specified above, and gauge how the maximum transmitted power and the resonant wavelength change due to nonlocality. The results computed for the mid IR regime, shown in Figs. 5-9a-5-9c differ from those in low THz frequencies, since we can observe that the relative blueshift $\Delta\lambda^* = \lambda^*_{LRA} - \lambda^*_{HM}$ is no longer independent of the diameter of the ring, with larger $\Delta\lambda^*$ reported for smaller diameters. There are two separate factors that should be mentioned: (1) at the FP0 resonance, the decay in relative blueshift as diameter increases is accentuated for the 1 nm; and (2) FP1 resonances across diameters are less sensitive to nonlocality compared to FP0, for both gaps. A possible explanation of this behavior relies on the permittivity function of alumina, which is frequency-dependent and becomes imaginary around $\lambda = 5$ µm, recall Fig. 5-7a. Hence, the alumina may become lossy for the FP0 reonances, since they are excited at longer wavelengths over broader ranges, compared to the more localized FP1 modes that occur at shorter wavelengths. This circumstance may also explain the larger $\Delta\lambda^*$ decay of FP0 1 nm compared to FP0 2 nm for increasing diameter. Conversely, the increase in transmitted power $\Sigma^* = \varsigma^*_{HM}/\varsigma^*_{LRA}$ due to nonlocality is fairly constant for FP0 resonances across diameter sizes, whereas for FP1 resonances it reduces with decreasing ring diameter. The most remarkable aspect is that, for 1 nm gap, the transmitted power with the HM increases between 60% and 80% compared to the LRA simulations.

5.2.3 Concluding remarks

In this section, we have presented a parametric study of annular nanogap structures for the far and the mid IR frequency regimes. The main takeaway of the study presented above is that one can exploit the construction of ROM within the HDG method to extract valuable and often unintuitive information that can guide the design process. To that end, we have shown the influence of the ring diameter and array periodicity on the position of the resonances, and obtained analytical expressions that relate these quantities with the resonant properties of the structure.

Moreover, we have demonstrated that the ROM framework allows us to assess, for instance, the impact of radius variations on the position of the resonance and the extraordinary optical transmission properties of the device. These studies can be conducted for any parameter that is subject to variability by naturally including it the ROM, which enables an efficient postprocessing.

Finally, we have completed the study of annular structures incorporating the effect of nonlocality, demonstrating that even in the far IR, we observe relative blueshifts of around 20% for 1 nm gap structures.

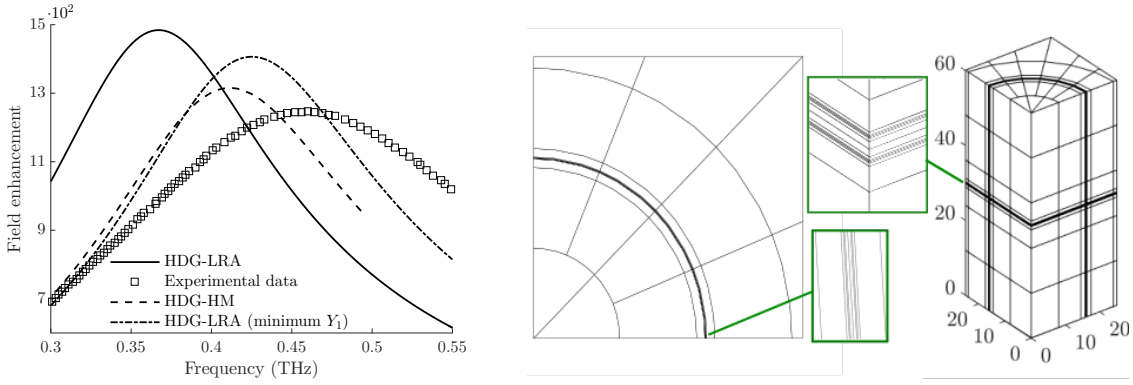
5.3 Experimental validation

The objective of this section is to validate the simulation results with experimental data reported in literature. The idea driving this study is twofold: on the one hand, assess the qualitative improvement obtained by introducing nonlocality in the simulations; on the other hand, use reduced order models to quantify and explain the mismatch observed between experiments and simulations [198, 277]. Even though there may be multiple sources of experimental discrepancy, we shall only focus on optical constants and basic geometric properties, that is gap size and radius. The optical constants are determined experimentally and thus subject to errors, whereas the extreme manufacturing constraints typically translate into structures whose geometry differs from the specified nominal values. Both measurement uncertainties and fabrication constraints can naturally be incoporated in the ROM framework.

5.3.1 Periodic annular structure at far infrared

First, we revisit the periodic structure in Fig. 3-4, and consider a gapsize of 2 nm. The experimental data for this structure was reported in [198], where a significant discrepancy was observed for the frequency-field enhancement profile using LRA, see Fig. 5-10a. The optical constants used for the LRA simulations are $\hbar\overline{\omega}_p = 9.02$ eV, $\hbar\overline{\gamma} = 0.02678$ eV and $\varepsilon_{Al_2O_3} = 3$.

In order to explain the mismatch, we first resort to nonlocality, and pursue simulations for the 2 nm nanogap with the HDG method for the hydrodynamic model. Since we are only interested in the value of the field enhancement, we may use a mesh of 1.1K cubic hexahedral anisotropic elements, which already gives a relative error less than 1% compared to the same mesh with quartic elements. The reduced problem for the traces comprises 138K unknowns, as opposed to the 215K unknowns had we used quartic elements. In Fig. 5-10b a 2d slice of the mesh for constant z is shown, with the inset showing the concentration of elements in the vicinity of the gap. In addition, we also present the entire 3d mesh, along with an inset that zooms in the gold film region. This highly anisotropic mesh allows us to solve for the full 3d EM wave field using a reduced number of degrees of freedom.



(a) (b)

Figure 5-10: (a) Field enhancement curve for the LRA, the HM and the LRA with minimum mismatch compared with experimental data. (b) 3d high-order mesh and 2d slice (with details) used in calculations.

The frequency sweep is depicted in Fig. 5-10a for the LRA and the HM, together with the experimental data. The inclusion of the hydrodynamic current produces a 13% blueshift in the resonance that partially bridges the gap between the LRA simulation results and experimental data. In addition, we also observe a decrease in the field enhancement, consistent

with both experiments and hydrodynamic theory. Thus, we conclude that a significant part of the experimental mismatch is due to inaccuracies of the physical representation of the phenomena, in this case the overly simplistic Drude model, that neglects nonlocal electron interactions at the nanoscale.

Parameter	Values	Frequency range [THz]	References
$\hbar\overline{\omega}_p$ [eV]	9.02 ± 0.18	< 12	[193, 197, 208]
$\hbar\overline{\gamma}$ [eV]	0.02678 ± 0.007	< 12	[193, 197, 208]
$\varepsilon_{Al_2O_3}$	2.8 - 3.2	thickness-dependent	[92]
n_{SiO_2}	1.96 ± 0.01	< 2	[166]

Table 5.3: Variability ranges for parameters, interval of validity and references for 2 nm alumina gap at low THz frequencies.

We now examine the impact of optical constants and geometry on the experimental mismatch for the LRA. The range of variation considered for each material parameter is given by literature, and is summarized in Table 5.3 for the 2 nm gap. For the geometric variables we prescribe a $\pm 5\%$ variation for both the gap width and the radius of the ring, typical for these structures. A ROM is developed for these six parameters, in addition to the frequency, using 800 snapshots that give rise to a 54 mode POD basis.

The impact of the aforementioned parameters is studied using global sensitivity analysis (GSA) via Sobol indices [228, 241]. In a nutshell, for each parameter X_i, GSA assesses its influence on a certain QoI $Y = f(\boldsymbol{X}_i)$ using two sensitivity indices

$$S_i = \frac{V_{X_i}(E_{\boldsymbol{X}_{-i}}(Y|X_i))}{V(Y)}, \qquad S_{Ti} = \frac{E_{\boldsymbol{X}_{-i}}(V_{X_i}(Y|\boldsymbol{X}_{-i}))}{V(Y)} \tag{5.1}$$

where $\boldsymbol{X}_{-i}$ refers to all parameters except X_i. The first order index S_i measures the expected reduction in variance of Y if X_i is fixed, whereas the total effect S_{Ti} quantifies the expected variance remaining if only X_i varies. That is, it measures the combined first and higher-order interactions of X_i with the other parameters. Note that $0 \leq S_i \leq S_{Ti} \leq 1$, and that $\sum_i S_i \leq 1$, where the equality only holds if the model is purely additive–there are no interactions, $S_i = S_{Ti}$. In order to compute numerical estimates of the indices in (5.1) we resort to the quasi-Monte Carlo variance-based expressions given by Saltelli *et al.* [227].

For this particular case, we consider as QoI the absolute blueshift $Y_1 = \nu^*_{data} - \nu^*$ in THz and the ratio of maximum field enhancement $Y_2 = \pi^*/\pi^*_{data}$. Hence, for each configuration of parameters an entire frequency sweep is required to identify the resonant frequency-

maximum field enhancement values (ν^*, π^*). The sensitivity indices are calculated with 40K quasi-Monte Carlo samples [227], and bootstrapping techniques [8] are used to compute error bars for the sensitivity indices. These are calculated with 1000 replicas of 40K resamples (with replacement) for a 95% confidence interval.

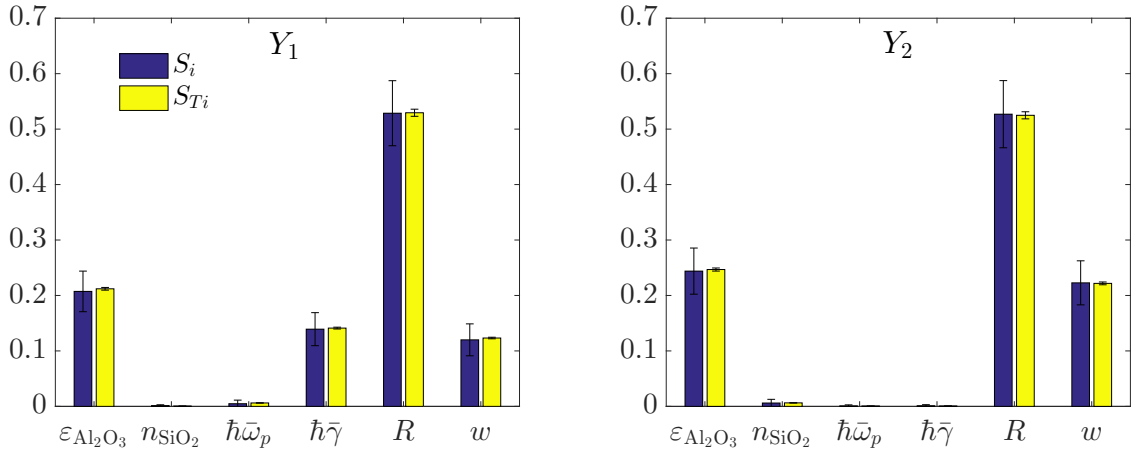


Figure 5-11: First order and total sensitivity indices with corresponding error bars of QoI Y_1 and Y_2 for 2 nm annular gap at FIR.

The sensitivity indices are reported in the bar chart 5-11. Both the permittivity of the substrate and the plasma frequency may be considered irrelevant to the variance of both QoI, whereas the collision rate only impacts the blueshift. All in all, the radius is the parameter that has the greatest impact on QoI, followed by the Al_2O_3 permittivity and the gap size. Moreover, despite the fact that confidence intervals for the first order indices are large, one can conclude that the model is almost additive, since the sum of total effect indices S_{Ti} is only marginally greater than one, with much sharper error bars. In absolute terms, the combination of parameters that minimizes Y_1 is given by $\hbar\bar{\gamma} = 0.01978$ eV, $\varepsilon_{Al_2O_3} = 2.8$, $R = 15.2$ μm and $w = 0.0021$ nm shown in Fig. 5-10a as dash-dot. This combination gives a relative blueshift of 16% with respect to LRA simulations with the original parameters.

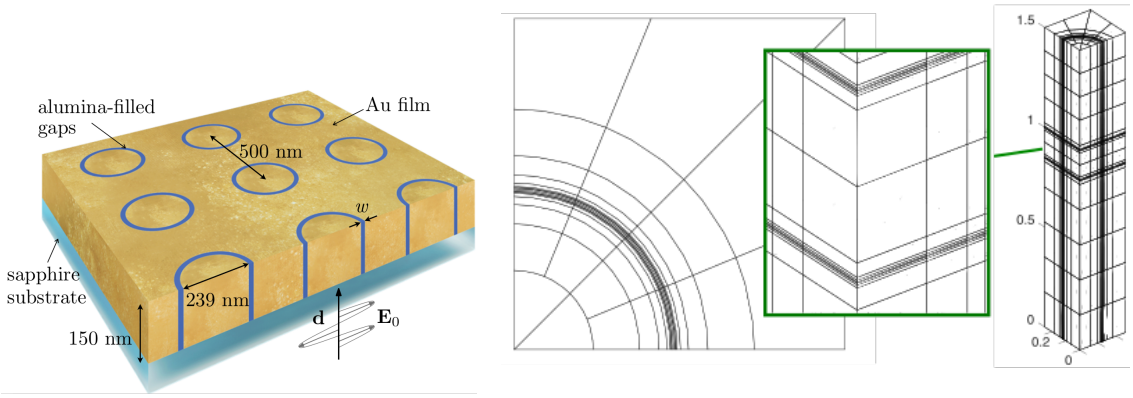
5.3.2 Periodic annular structure at mid infrared

For the second example we focus on a periodic annular structures for frequencies in the mid IR, for which experimental data is available [277] for several gapsizes. The structure under consideration is an array of 125 nm radius rings with a 500 nm periodicity, and we shall focus on the 1 and 2 nm gap width. The gap is fabricated with Al_2O_3, whose permittivity in

the mid IR is shown in Fig. 5-7a, whereas the substrate used was sapphire. The refractive index of sapphire [59] is given by the following Sellmeier formula

$$n^2_{\text{sapphire}} - 1 = \frac{1.4313493\lambda^2}{\lambda^2 - 0.0726631^2} + \frac{0.65054713\lambda^2}{\lambda^2 - 0.1193242^2} + \frac{5.3414021\lambda^2}{\lambda^2 - 18.028251^2} \tag{5.2}$$

where λ is in microns. Sapphire becomes lossy for wavelengths greater than 9 μm, but for the wavelengths under consideration, we assume its refractive index is real and given by the formula (5.2).



(a) (b)

Figure 5-12: (a) Schematic of periodic annular structure. (b) 3d high-order mesh for $w = 2.3$ nm and 2d slice (with details) used in calculations.

The excitation of resonances for ring structures at mid infrared frequencies requires array periodicites in the order of nanometers. Hence, manufacturing is specially challenging, which often translates into discrepancies between the nominal and measured specifications. Consequently, for the numerical simulations we employ the geometric features reported in the Supplementary Information of [277]. These correspond to a ring radius of 119.5 nm and gap apertures w of 1.2 and 2.3 nm respectively, sketched in Fig. 5-12a. The difference between the nominal and the measured geometric features is considerable – around 20% for the 1 nm gap–, thus stressing the importance of incorporating geometry imperfections in the simulations.

Despite these adjustments for geometry, the reported LRA simulations in [277] show a significant mismatch compared to experimental data, see Figs. 5-13a and 5-13c. The optical constants for gold used in the simulations are $\hbar\overline{\omega}_p = 8.45$ eV, $\hbar\overline{\gamma} = 0.047$ eV, whereas the dielectric constants for sapphire and alumina are given by expression (5.2) and Fig.

Parameter	Values	Wavelength range [μm]	References
$\hbar\overline{\omega}_p$ [eV]	8.45 ± 0.44	0.3-25	[190]
$\hbar\overline{\gamma}$ [eV]	0.047+[-0.0083,0.0128]	0.3-25	[190]
$\varepsilon_{Al_2O_3}$	Fig. 5-7a	0.3-14.3	[22, 128]
n_{sapphire}	Equation (5.2)	0.2-9	[59]

Table 5.4: Variability ranges for parameters, interval of validity and references at mid IR frequencies.

5-7a respectively. As before, we explore two separate avenues to explain the experimental discrepancy, that is the hydrodynamic model for electron motion and the impact of uncertain optical constants, whose variability is described in Table 5.4, and geometry.

Again, the hydrodynamic simulations are carried out on the cubic mesh shown in Fig. 5-12b, using the nominal value of the parameters, and we focus only on the zeroth Fabry-Pérot mode for which the enhancement is constant along the aperture. The simulation results for both the LRA and the HM are reported in Figs. 5-13a and 5-13c, where we observe the significant discrepancy between the LRA and experimental data is partially explained with the nonlocal effects. The relative blueshift in wavelength is 11% for the 1.2 nm gap and 7% for the 2.3 nm gap, with a relative increase in transmission of 60% and 27% respectively.

Alternatively, we combine ROM with GSA to quantify the influence of the Drude model parameters in Table 5.4 on the experimental mismatch, along with a prescribed ±5% relative variation on the gap and the radius. The ROM is developed for these 4 parameters and the frequency using 500 snapshots, and resulting in a basis of 45 (resp. 38) modes for 2.3 (resp. 1.2) nm gap. The quantities of interest that will be queried for GSA are the absolute blueshift $Y_1 = \lambda^* - \lambda^*_{data}$ in microns and the ratio of maximum transmitted power $Y_2 = \varsigma^*/\varsigma^*_{data}$, and follow the same procedure as in the far IR case to compute the sensitivity indices and its errorbars.

The first conclusion drawn from the bar charts in 5-13b (2.3 nm) and 5-13d (1.2 nm) is that the interactions between parameters are negligible. The blueshift may be mostly explained by the prescribed radius variation, whereas the increase in transmission is basically due to the uncertainty in the collision rate $\overline{\gamma}$. Note that the influence of the Drude model parameters on the outputs of interest is different to the far IR, where the plasma frequency is irrelevant and the collision rate only affects the shift. Indeed, the behavior of metals at far IR modeled by Drude permittivity approaches that of perfect conductors. Conversely, in the mid IR regime ohmic losses are significantly lower, allowing a greater penetration of

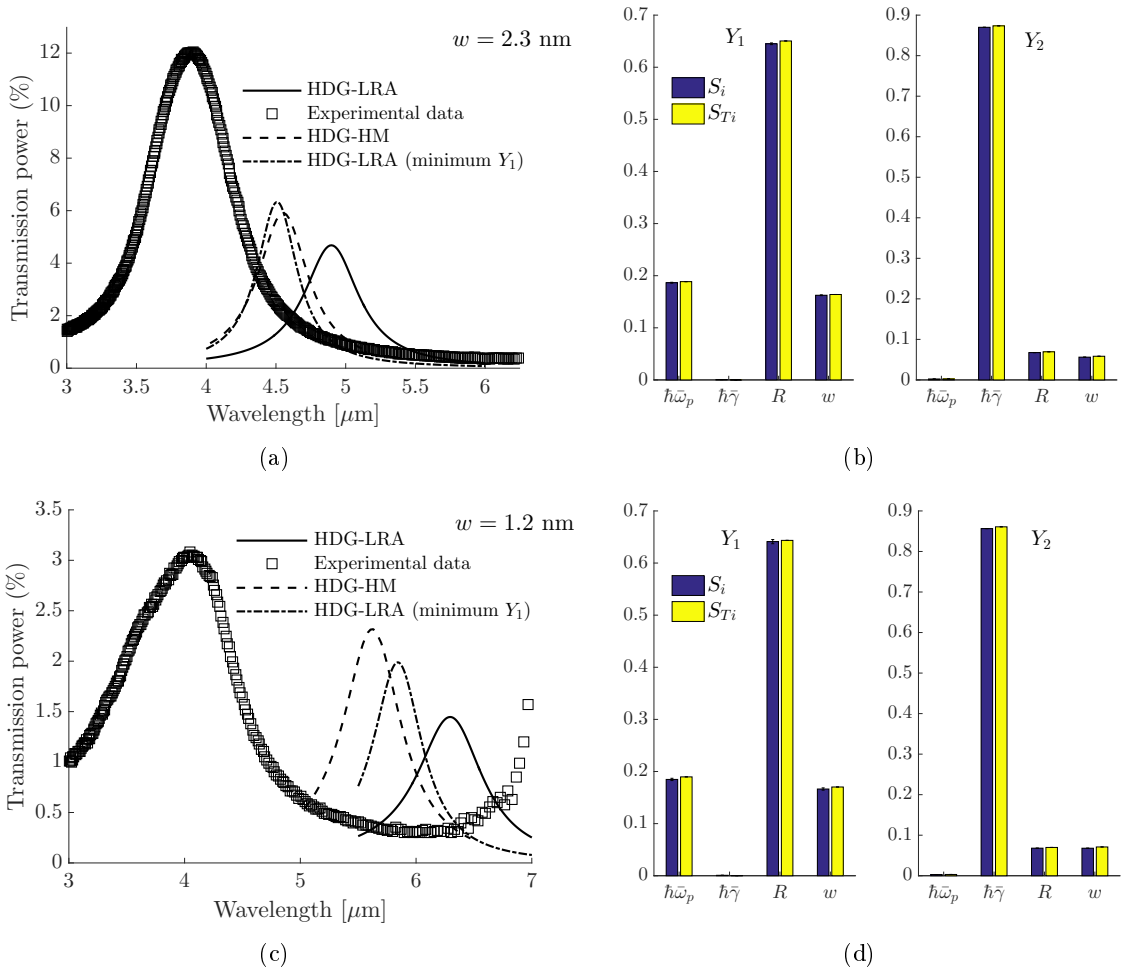


Figure 5-13: Field enhancement curve for the LRA, the HM and the LRA with minimum mismatch parameters compared with experimental data for (a) 2.3 nm gap and (c) 1.2 nm gap. First order and total sensitivity indices with corresponding error bars of QoI Y_1 and Y_2 for (b) 2.3 nm gap and (d) 1.2 nm gap at mid infrared.

the incident light in the metal. Hence, their effect on the resonances is entirely different. For both gap widths, the parameter combination that minimizes Y_1 is $\hbar\overline{\gamma} = 0.0387$ eV, $\hbar\overline{\omega}_p = 8.89$ eV, $R = 113.525$ nm and $w = 0.00241$ nm (resp. $w = 0.00126$ nm), and the transmission for these values of the parameters is shown in Figs. 5-13a (resp. 5-13c) as dash-dot. This combination gives a relative blueshift of 7% for both gap apertures, which is comparable to the blueshift caused by nonlocality for 2.3 nm, but inferior to the blueshift for the 1.2 nm gap.

5.3.3 Concluding remarks

To sum up, in the process of validating simulation results with experimental data in plasmonic applications, it is of paramount importance to not only model appropriately the behavior of electrons in metals, but also to account for variability in the parameters arising from manufacturing or measurement errors. We have shown that both avenues can be important in explaining the discrepancy for annular nanogaps at different IR regimes. Nonetheless, we have not been able to fully explain the experimental mismatch with our models, which is specially severe as the gap shrinks to nanometric sizes.

The important conclusion that should be drawn from the GSA is that an accurate geometric characterization is crucial, since minor modifications greatly impact the numerical results. Furthermore, the fact that the model is almost additive opens the possibility of developing, instead of a single ROM with multiple parameters, a collection of ROMs with a single parameter in addition to the frequency. Reduced order modeling techniques are more effective for a smaller number of parameters, due to the curse of dimensionality. The major advantages of neglecting interactions are not only the inferior number of parameter values required to exhaustively explore the parameter space, but also the reduced dimension of the orthonormal bases after compression, thus accelerating the online evaluation.

Furthermore, only gap and radius modifications in the radial direction have been considered, although annular rings often present defects in the circumferential direction as a consequence of the manufacturing constraints [277]. Hence, a natural extension of the methodology presented here would be to incorporate more complex geometry variations in the ROM framework. The main drawback is the significant computational cost it entails, since it requires not only a finer spatial discretization to represent the defects, but also complicates the empirical interpolation necessary for an efficient ROM implementation.

In terms of physical modeling, the immediate future step is to augment the physics by incorporating quantum effects such as quantum electron tunnelling. These effects, deliberately omitted by both the LRA and the HM, are however present for nanometer-wide apertures. We therefore anticipate a better agreement between simulations and experiments in those cases.

5.4 Design of a concentric ring structure

The applications involving annular structures considered in this thesis have only accounted for a single nano-aperture. Nonetheless, structures consisting of multiple concentric 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 naturally arises, as different configurations may lead to dramatic changes in transmission. In this section, 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.

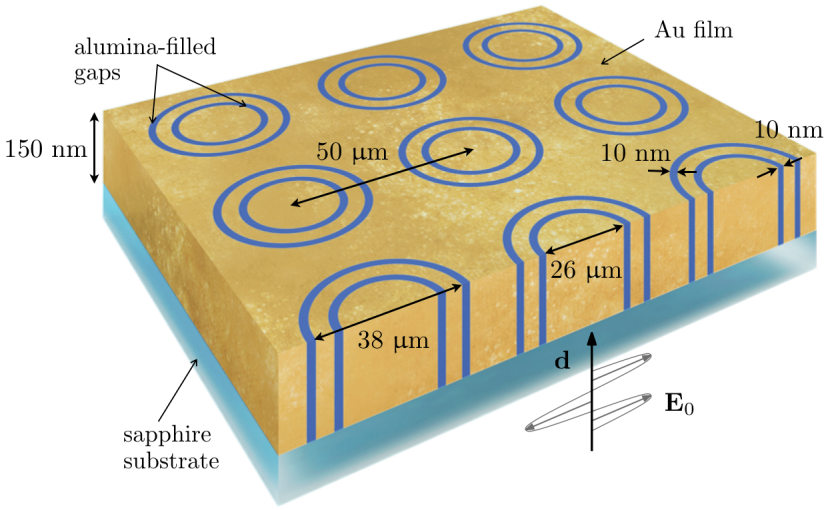


Figure 5-14: Schematic diagram of thin gold film on sapphire substrate patterned with periodic square array of concentric alumina gaps under plane wave THz illumination.

The reference structure is shown in Fig. 5-14, with ring diameters of 26 and 38 microns. The material parameters and gold optical constants are set to $\varepsilon_{\mathrm{Al_2O_3}} = 5.5$, $n_{\mathrm{sapphire}} = 3.07$, $\hbar\omega_p = 9.02$ eV, $\hbar\gamma = 0.02678$ eV, $\varepsilon_\infty = 1$, and we consider frequencies in the range 0.3 to 0.9 THz. The gap widths are fixed to 10 nm, thus the only tunable geometric parameters are the radii of the rings, 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. The deformation mapping required to simultaneously deform two concentric rings is a straightforward extension of that described in Appendix B. We then resort to the HDG-POD method for Maxwell's equations on a reference domain to construct a surrogate of the high-fidelity model, using 650 snapshots computed over the 3-dimensional parametric space formed by the frequency and the radii. The resulting ROM consists of 74

modes, and its accuracy is examined in Fig. 5-15 evaluating the relative error for a test set of 100 parameter combinations.

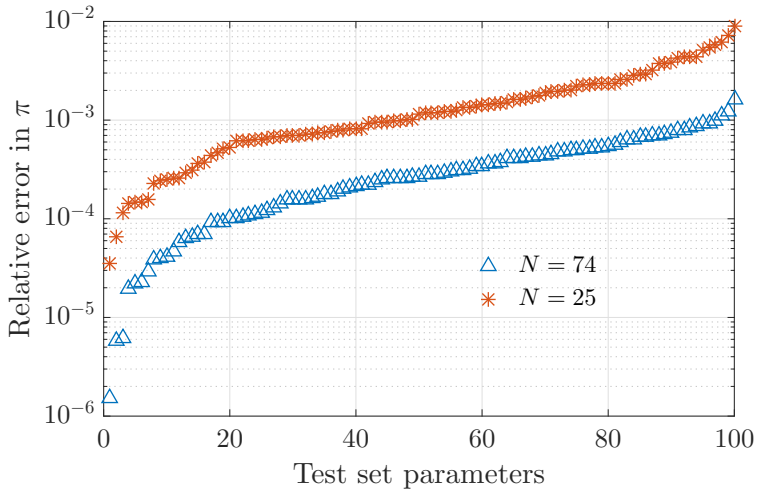


Figure 5-15: Relative error in field enhancement for test set parameters, evaluated for 74 and 25 POD modes.

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

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

where E_ν, V_ν refer to the expectation and variance respectively, for frequencies in the range 0.4 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 and compute the objective function invoking the ROM for all possible combinations of radii, thus avoiding the use of optimization algorithms.

In Fig. 5-16a we show the objective functions above evaluated at multiple values of the parameters, normalized between 0 and 1 for simplicity, along with the maximizing parameter configuration marked with a cross. For each objective function, we depict in Fig. 5-16b the transmitted power ς spectrum for the optima radii (solid line), and compare it with the power transmitted by a structure with a single ring, for both the inner (dashed) and the outer (dash-dot) radius. 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. 5-16a (left) suggests that larger transmissions may be encountered if the separation is further reduced.

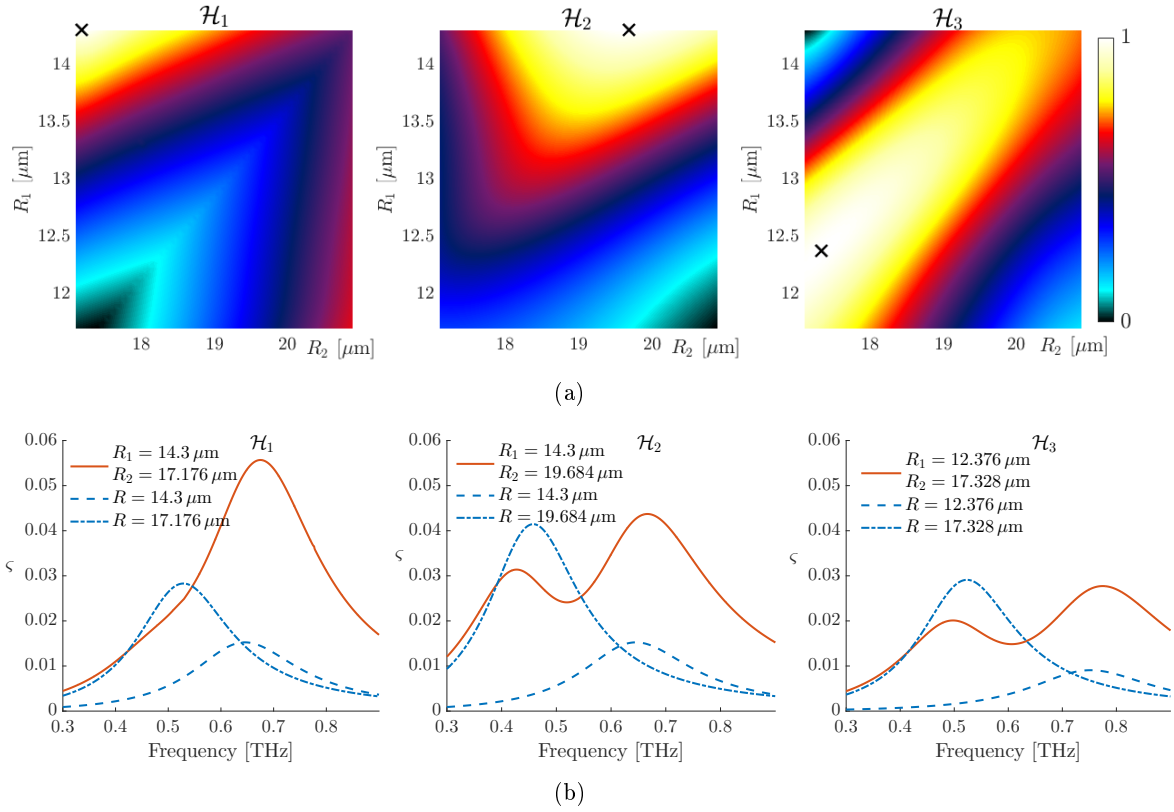


Figure 5-16: (a) Objective functions shown as surface plot for innner R_1 and outer R_2 radii values, normalized to [0, 1]. Optimal configuration is shown with black cross. (b) Frequency-transmission profiles for optimal configurations.

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.

The main takeaway of this study is the demonstration of how reduced order modeling techniques can be applied to the design of plasmonic devices, since they enable a fast evaluation

of the objective function. Despite focusing on the two ring case, the extension to multiple concentric rings is natural, and may potentially lead to superior designs. The main caveat when accounting for more design parameters is always the construction of the ROM, since a greater number of snapshots is typically required to better explore the parameter space. In addition, multiple rings imply finer HDG discretizations, therefore increasing computational cost to obtain each snapshot. However, once the surrogate model has been built, optimization can be performed efficiently. Even the computation of gradients, necessary to expedite the design process, does not entail an excessive computational burden when the ROM has been constructed.

Chapter 6

Conclusions and future work

We conclude with the main contributions and future research directions of this thesis.

6.1 Summary of contributions

The overarching goal of this thesis has been the development of a computational framework to simulate electromagnetic wave propagation problems for metallic nanostructures. We have extended a state-of-the-art high fidelity numerical scheme, the hybridizable discontinuous Galerkin method, for two models of the electron motion in metals. We have also developed a reduced order modeling capability to efficiently address parametric variations.

The HDG method for metals using the Drude's permittivity is able to accommodate an enormous disparity of length scales, accurately represent complex geometries and material discontinuities at interfaces and capture tightly localized fields. This numerical scheme enables the efficient simulation of Maxwell's equations for frequency regimes and structures beyond the scope of other existing approaches, such as COMSOL Multiphysics [1] or Lumerical [2], thus constituting a powerful method for simulation and design of plasmonic structures.

The HDG formulation and implementation of the hydrodynamic model for noble metals involves the extension of the HDG approach to grad-div operators. Consistency and well-posedness of the weak formulation of HDG method for the hydrodynamic model have been provided. The HM involves solving an extra PDE in the metallic region, requiring a coupling

condition at the metal-dielectric interface. In terms of computational challenges, besides the obvious increase in degrees of freedom that stems from the additional equation, the physics modeled by the hydrodynamic pressure are more complex, and significantly harder to capture for the numerical scheme. Indeed, the nonlocality generates a boundary-layer structure inside the metal, as a consequence of the spreading of the electron density distribution at the metal boundaries. Properly resolving these deep-subwavelength features requires finer discretizations, that along with the additional degrees of freedom constitute a significant computational barrier. The proposed implementation leverages common HDG advantages, such as high-order accuracy and local elimination of degrees of freedom, and is able to overcome the aforementioned limitations. In terms of predictive power, the hydrodynamic model produces a better agreement with experimental results for noble metal nanostructures than the Drude model.

Furthermore, we have developed a reduced order modeling capability for the HDG method with the local response approximation. To that end, we first introduced two new weak formulations for the HDG method that enable the efficient treatment of material parameters, such as optical constants of dielectrics and metals, as well as geometric parameters that model deformations of the structure. We then combine these weak formulations with interpolation and model reduction techniques, such as discrete empirical interpolation and proper orthogonal decomposition. The result are surrogate models that provide accurate yet inexpensive simulations of the desired plasmonic structure. The use of an offline-online computational strategy renders this approach particulary useful for scenarios where solutions for multiple parameter values are required. The accurate characterization of optical properties and geometry specifications at the nanoscale can be difficult, and has a severe impact on the fidelity of the simulations. The proposed methodology attempts to alleviate these limitations. A main contribution has been to fuse model reduction techniques with the HDG method adapted for metallic nanostructures, leading to enhanced numerical prediction capabilities for both simulation and design.

Finally, we have demonstrated the performance of the above methodologies for various realistic 3d problems and applications in the realm of periodic gap structures. The applications involve a study of a 2d nanoslit, a parametric analysis of annular gap structures, comparisons between simulations and experimental data and the design of a periodic concentric

ring. Indeed, the methodology developed in this thesis has enabled us to tackle problems and regimes that had not been addressed before, paving the way towards a deeper understanding of plasmonic devices and their behavior.

6.2 Future work

We now describe some possible lines of future work, together with the challenges and potential benefits for the simulation and design of plasmonic devices.

6.2.1 Incorporation of quantum effects

The simulation of metallic nanostructures developed in this thesis has considered two different models to characterize the motion of electrons, namely the Drude permittivity and the hydrodynamic model. We have extensively shown that the hydrodynamic model, albeit more computationally demanding, produces results which are in better agreement with experimental data in the nanometric regime for noble metals. Nonetheless, these models omit effects such as quantum electron tunnelling, which is known to greatly impact the performance of nanoparticles and dimers at subnanometric scales [69,229,233,249].

Current state-of-the-art quantum models were already discussed in Section 2.5.6, and the most prominent approaches [44, 69, 70, 251] have already been described in terms of PDE systems. The challenge is not only to derive an HDG implementation for the quantum models, but also to devise an efficient hybridized formulation that minimizes the number of globally coupled degrees of freedom.

In [69, 70], Esteban *et al.* proposed to include quantum effects by prescribing a special permittivity for the gap region, which models the tunnelling current density between the metallic surfaces. Alternatively, the approaches proposed in [44, 251] require solving additional equations. As opposed to the hydrodynamic model, where the additional equations are only prescribed within the metal, these latter quantum models establish a system of augmented PDEs in the metal and its surroundings. This is a consequence of tunnelling effects, which cause the electrons to escape the metal. Mathematically, regions where electrons flow are dictated by a nonzero equilibrium electron density $n_0(\overline{\mathbf{x}}) > 0$. The complication stems

from the fact that $n_0(\overline{\mathbf{x}})$ is not known *a priori*, and thus needs to be computed either within the simulation by solving a nonlinear equation [251] or specified beforehand [44]. The latter is only available for certain simple structures such as spheres. In any case, the size of the global linear system will perforce increase, thus adding further computational complexity.

It would therefore be interesting to account for quantum effects within the HDG method. This would endow the numerical methods presented here not only with an enhanced predictive power, but also with the ability to simulate quantum-controlled devices within the exciting emerging field of quantum plasmonics [248].

6.2.2 Numerical simulation

Consideration of embedded DG schemes

Numerically obtaining the electromagnetic wave solutions for extended 3d structures is computationally demanding, despite the use of advanced numerical schemes like the hybridizable discontinuous Galerkin method. The HDG method requires the solution of a linear system for the globally coupled degrees of freedom defined on the faces of the discretization, which are of discontinuous nature. Hence, the duplication in degrees of freedom –at the nodes in 2d, along edges in 3d– significantly contributes to the size of the linear system to be solved.

This duplication motivated the development of the class of embedded discontinuous Galerkin (EDG) methods [51,180,200], which enforce a continuous approximation space for the hybrid unknowns, thus leading to systems of equations of much smaller size. In fact, the systems arising from EDG discretizations present the sparsity patterns of continuous Galerkin methods with static condensation. The main drawback of EDG methods are the loss of one order of convergence of the fluxes due to the single-valuedness of the hybrid variable [51] compared to the standard HDG method. In addition, one loses the natural treatment of boundary conditions provided by the HDG method, specially in boundary edges where different conditions need to be prescribed for each adjacent boundary face. A promising fix for this issue is the newly proposed class of interior EDG (IEDG) methods [74, 180], whereby the approximation space of hybrid variables is continuous in the interior and discontinuous in the boundaries. The IEDG leads to systems with dimension slightly larger than those of EDG, but significantly smaller than those of HDG, while accurately enforcing the boundary

conditions. Hence, IEDG is an attractive alternative to HDG methods if computational power and storage is the limiting factor.

Nonetheless, implementing Maxwell's equations with IEDG precludes the possibility of using the tangent vectors to approximate the solution on the traces, since they are not uniquely defined at the edges. Hence, in 3d with IEDG we need to solve for the three components of the electrical field's trace, as opposed to two components for HDG. The computational savings of IEDG will therefore be more modest for Maxwell's equations.

Implementation of more efficient radiation conditions

Simulating unbounded domains is a well known challenge for computational electromagnetics with finite element methods. In this thesis, we have only employed first-order absorbing boundary conditions as a means to simulate infinite domains, since they can be readily implemented without adversely affecting the computational cost and produce satisfying results for normally incident waves. A better approach are high-order absorbing boundary conditions, see [84] for a review, that ensure greater absorption at the expense of locality and convenience of implementation.

A widespread alternative is the use of PMLs [17, 95, 254], which were already discussed in Chapter 1. PMLs consist of additional layers that cause an exponential decay of the outgoing waves, impeding reflection at the boundaries. The main shortcomings of PMLs are the increased computational requirements, since the domain of interest needs to be augmented with the layers of absorbing materials, and the choice of the parameters modeling the decay of the outgoing waves, which usually involves fine tuning if one is interested in minimizing the size of the extended domain.

Incorporating either high-order ABCs or PMLs to the computational framework presented in this thesis should be addressed. Recent studies show similar numerical effectivity from both alternatives [207], which in any case lead to a superior performance compared to the first-order absorbing boundary conditions used in this thesis.

Use of iterative solvers

In this thesis, only the direct solver implemented in MATLAB has been employed for linear systems. Furthermore, the computations have been mainly carried out on a 512GB `Linux` machine with four eight-core machine with 32 AMD Opteron(tm) processors 6320x15, which have allowed for a parallel computation of the snapshots required to form the several reduced order models. Nonetheless, the size of the linear systems in 3d is a bottleneck for direct solvers. Unfortunately, the poor scalability of the linear system in terms of degrees of freedom cannot be circumvented by using more powerful machines.

Hence, an option that should be considered in the future is the use of iterative methods. Combined with efficient preconditioners, they would enable the simulation of much larger problems, albeit increasing the computational time. Furthermore, a parallel implementation of iterative solvers is required to achieve superior performance, as it has already been demonstrated for the HDG discretization of Navier-Stokes equations [74,217]. The main drawback for time-harmonic Maxwell's equations is the lack of robust preconditioners. Consequently, it may be necessary to solve the equations in the time domain formulation if iterative solvers are essential to compute large scale solutions.

6.2.3 Reduced order modeling

Error bounds

The main shortcoming of the ROM strategy presented in this thesis is the lack of *a posteriori* error bounds to quantify the error incurred by the surrogate. In addition, the definition of inexpensive and sharp error bounds may further benefit the snapshot computation, since they may be used to drive the sampling process in a greedy fashion [86, 88, 257, 259]. In short, instead of computing the snapshots for a preestablished set of parameters, we may sequentially select the input parameters with the largest error bound, and iteratively enrich the model.

For noncoercive problems such as time-harmonic Maxwell's equations error bounds have been developed using the successive constraint method [110, 113], where a strategy to efficiently compute approximations to the inf-sup constant is devised. More recently, a novel approach

based on a statistically modeling the mapping between inexpensive error bounds and the actual error using Gaussian processes has been proposed [60]. Alternatively, error bounds may also be constructed using the mapping between input parameters and the error of the surrogate. Even though this mapping is not directly computable, machine learning techniques such as deep learning [162] arise as valid methodologies to learn the mapping from computational data. The advantage of neural networks is that no assumptions on the regularity and conditions of the data are needed, thus being ideal for complex problems such as the one under consideration, with geometry interpolation and arbitrary outputs.

Hydrodynamic model

The ROM framework has only been developed for the time-harmonic Maxwell's equations with Drude metals, enabling parametric studies and assessment of uncertainty in both the material properties and the geometric specifications. Thus, a natural extension is seeking a dimensionality reduction strategy for the hydrodynamic model for metals.

Nonetheless, further complications can be anticipated. Firstly, the computation of a snapshot collection prior to compression may be very demanding for the hydrodynamic model. As we have repeatedly shown throughout this thesis, the number of unknowns (specially in 3d), combined with the tightly localized structure of the EM solution negatively impacts the computational requirements for the direct solution of the linear system.

Secondly, it is unclear that a reduced order model developed globally will be able to accurately capture the subnanometric boundary layer that arises at the metal-dielectric interface, a consequence of the hydrodynamic pressure term. It may therefore be interesting to combine standard ROM techniques with domain decomposition approaches, which enable the construction of surrogates for multiscale problems [41, 111, 152, 263].

6.2.4 Applications

Simulation

The focus of this dissertation has been on periodic square arrays of annular gap structures. Beyond periodic arrays, there is a myriad of relevant plasmonic structures that would benefit

from the work presented in this thesis. For instance, nanoparticles are being used nowadays for a broad range biomedical applications such as drug delivery, biosensing, disease detection and cancer theranostics [7,27,33,58,62,66,104–106,124,133,141,264]. Nanoparticles are attractive because they can be manufactured in multiple sizes and shapes, such as spheroids, rods, stars, cubes, cages or shells, and exhibit a high tunability of their optical properties. Hence, it would be interesting to combine HDG and ROM techniques to provide not only their optical response, but also its dependence on geometry modifications over several frequency regimes.

Another example of plasmonic structures of interest are plasmonic antennas for advanced data storage [35,156,187,245,280], where the extreme confinement and enhancement capabilities of nanoantennas are leveraged towards a high-efficient recording by locally heating the magnetic medium, known as heat-assisted magnetic recording.

Design

In this thesis, we have demonstrated how the construction of ROM may be further used to design plasmonic devices, in this case a concentric ring structure. This methodology can be readily applied to the simulation and optimization of plasmonic devices with multiple geometry design variables, for example annular apertures with several concentric rings. Another potential application are bullseye grating structures [135] combined with annular nanogaps, where the depth, width and pitch of the grooves can be treated as design parameters.

On the other hand, an interesting although challenging avenue is producing designs that are robust under geometry modifications. For instance, one could consider small variations in the nominal radius and the gap, or even non-homogeneous variations in the radial direction. These problems are commonly cast as stochastic optimization problems. Consequently, after constructing the ROM, we need strategies to evaluate objective functions involving statistical quantities, perhaps defined over a multidimensional stochastic space. To that end, it may be useful to recall methodologies that combine model reduction with variance reduction techniques [261–263], enabling the fast computation of stochastic objective functions.

2d materials

A new and promising line of research is the combination of plasmonic devices with layers of 2d materials, such as graphene [278] or black phosphorus [137]. The most definitive feature is that high ohmic losses incurred by noble metals in plasmonic devices, for most frequency regimes, can be reduced by placing a single layer of doped two-dimensional atoms on top of the metallic structure. 2d materials exhibit remarkable tunability by electrical or chemical doping, enabling them to sustain plasmons for broad range of frequencies, with confinements in volumes several orders of magnitude smaller than the diffraction limit. All in all, 2d materials offer the possibility of dramatically altering their optical properties, as well as the behavior of noble metals, thus paving the path towards revolutionary applications in nanophotonics [89, 115, 123, 129, 144].

From the computational standpoint, simulating the interaction of noble metals with 2d materials is inherently hard. The disparity in length scales is even more severe than in conventional plasmonics, since we need to simulate atom-thin layers. In addition, graphene and black phosphorus exhibit a highly anisotropic behavior, which needs to be taken into account as a nondiagonal conductivity tensor. Nonetheless, we strongly believe that the high-fidelity numerical methods introduced in this thesis may be adapted and further improved to pursue simulation and robust design of ever complex plasmonic structures.

Appendix A

Properties of HDG for hydrodynamic model

In this appendix, we complete the definition of the HDG method for Maxwell's equation with the hydrodynamic model introduced in Chapter 3, proving it is locally conservative, consistent and well defined.

Proposition 1. *The HDG method defined by* (3.11) *is locally conservative and consistent.*

Proof. When ε_∞ is constant on each face, the last two equations of (3.11) imply that

$$\begin{aligned}\llbracket \mathbf{n} \times \widehat{\mathbf{V}}_h \rrbracket &= 0, \quad \text{on } \mathcal{E}_h^o,\\ \llbracket \mathbf{n} \cdot \widehat{\mathbf{J}}_h \rrbracket &= 0, \quad \text{on } \mathcal{E}_h^o.\end{aligned}$$

Substituting (3.5) and (3.10) into the expressions above we arrive to

$$\begin{aligned}\llbracket \mathbf{n} \times \mathbf{V}_h \rrbracket + \tau_t^+ \mathbf{E}_h^+ + \tau_t^- \mathbf{E}_h^- - (\tau_t^+ + \tau_t^-)\widehat{\mathbf{E}}_h &= 0, \quad \text{on } \mathcal{E}_h^o,\\ \llbracket \mathbf{n} \cdot \mathbf{J}_h \rrbracket - \tau_t^+ U_h^+ - \tau_t^- U_h^- + (\tau_t^+ + \tau_t^-)\widehat{U}_h &= 0, \quad \text{on } \mathcal{E}_h^o.\end{aligned}$$

Isolating the value of the traces we get

$$\begin{aligned}\widehat{\mathbf{E}}_h &= \frac{\tau_t^+ \mathbf{E}_h^+ + \tau_t^- \mathbf{E}_h^- + \llbracket \mathbf{n} \times \mathbf{V}_h \rrbracket}{\tau_t^+ + \tau_t^-}, \quad \text{on } \mathcal{E}_h^o,\\ \widehat{U}_h &= \frac{\tau_t^+ U_h^+ + \tau_t^- U_h^- - \llbracket \mathbf{n} \cdot \mathbf{J}_h \rrbracket}{\tau_t^+ + \tau_t^-}, \quad \text{on } \mathcal{E}_h^o.\end{aligned} \tag{A.1}$$

Substituting these expressions into (3.5) and (3.10) we obtain

$$\begin{aligned}\widehat{\mathbf{V}}_h &= \frac{\tau_t^+\mathbf{V}_h^- + \tau_t^-\mathbf{V}_h^+ + \tau_t^+\tau_t^-[\![\mathbf{E}_h\times\mathbf{n}]\!]}{\tau_t^+ + \tau_t^-}, \quad \text{on } \mathcal{E}_h^o,\\ \widehat{\mathbf{J}}_h &= \frac{\tau_t^+\mathbf{J}_h^- + \tau_t^-\mathbf{J}_h^+ - \tau_t^+\tau_t^-[\![U_h\mathbf{n}]\!]}{\tau_t^+ + \tau_t^-}, \quad \text{on } \mathcal{E}_h^o.\end{aligned} \tag{A.2}$$

The expressions (A.1) and (A.2) show that the numerical traces of the HDG method are single valued across inter-element faces, hence the HDG method is locally conservative, by virtue of the definition of local conservation introduced in [10]. Furthermore, since $\mathbf{E}\in\boldsymbol{H}^{\text{curl}}(\overline{\Omega})$ and $U\in H^1(\overline{\Omega})$, we have $\widehat{\mathbf{E}}=\mathbf{E}^t$ and $\widehat{U}=U$ on $\mathcal{E}_h$. It follows from expressions (3.5) and (3.10) that $\widehat{\mathbf{V}}=\mathbf{V}$ and $\widehat{\mathbf{J}}=\mathbf{J}$. Finally, if we substitute them into the first four equations of (3.9) and integrate by parts again, we arrive to

$$\begin{aligned}(\mathbf{V}-\nabla\times\mathbf{E},\boldsymbol{\kappa})_{\mathcal{T}_h} &= 0,\\ (\beta^2\nabla U+\omega(\omega+i\gamma)\mathbf{J}-i\omega\omega_p^2\mathbf{E},\boldsymbol{\eta})_{\mathcal{T}_h} &= 0,\\ (\nabla\times\mathbf{V}-\omega^2\varepsilon_\infty\mathbf{E}-i\omega\mathbf{J},\boldsymbol{\xi})_{\mathcal{T}_h} &= 0,\\ (U-\nabla\cdot\mathbf{J},\zeta)_{\mathcal{T}_h} &= 0.\end{aligned}$$

The exact solution of (2.11) is therefore a solution of the HDG formulation (3.9), thus the HDG method is consistent. □

In addition, it can also be shown that the HDG method is well defined, that is there exists only one solution.

Proposition 2. *Assume that both $\omega^2\varepsilon_\infty$ and $\omega(\omega+i\gamma)$ are different from the eigenvalues λ_1, λ_2 of the following eigenproblem: find λ_1, $\lambda_2\in\mathbb{C}$ and $(\mathbf{N}_h,\mathbf{Q}_h,\mathbf{S}_h,\psi_h,\widehat{\mathbf{Q}}_h,\widehat{\psi}_h)\in\boldsymbol{W}_h\times\boldsymbol{W}_h\times\boldsymbol{W}_h\times W_h\times\boldsymbol{M}_h(\mathbf{0})\times M_h$ such that*

$$\begin{aligned}(\mathbf{N}_h,\boldsymbol{\kappa})_{\mathcal{T}_h}-(\mathbf{Q}_h,\nabla\times\boldsymbol{\kappa})_{\mathcal{T}_h}-\langle\widehat{\mathbf{Q}}_h,\boldsymbol{\kappa}\times\mathbf{n}\rangle_{\partial\mathcal{T}_h} &= 0,\\ -\beta^2(\psi_h,\nabla\cdot\boldsymbol{\eta})_{\mathcal{T}_h}+\beta^2\langle\widehat{\psi}_h,\boldsymbol{\eta}\cdot\mathbf{n}\rangle_{\partial\mathcal{T}_h}-i\omega\omega_p^2(\mathbf{Q}_h,\boldsymbol{\eta})_{\mathcal{T}_h} &= -\lambda_2(\mathbf{S}_h,\boldsymbol{\eta})_{\mathcal{T}_h},\\ (\nabla\times\mathbf{N}_h,\boldsymbol{\xi})_{\mathcal{T}_h}+\tau_t\langle\mathbf{Q}_h-\widehat{\mathbf{Q}}_h,\mathbf{n}\times\boldsymbol{\xi}\times\mathbf{n}\rangle_{\partial\mathcal{T}_h}-i\omega(\mathbf{S}_h,\boldsymbol{\xi})_{\mathcal{T}_h} &= \lambda_1(\mathbf{Q}_h,\boldsymbol{\xi})_{\mathcal{T}_h},\\ -(\nabla\cdot\mathbf{S}_h,\zeta)_{\mathcal{T}_h}+(\psi_h,\zeta)_{\mathcal{T}_h}+\tau_n\langle\psi_h,\zeta\rangle_{\partial\mathcal{T}_h}-\tau_n\langle\widehat{\psi}_h,\zeta\rangle_{\partial\mathcal{T}_h} &= 0,\\ -\langle\mathbf{n}\times\mathbf{N}_h+\tau_t(\mathbf{Q}_h-\widehat{\mathbf{Q}}_h),\boldsymbol{\mu}\rangle_{\partial\mathcal{T}_h} &= 0,\\ \langle\mathbf{S}_h\cdot\mathbf{n},\theta\rangle_{\partial\mathcal{T}_h}-\tau_n\langle\psi_h,\theta\rangle_{\partial\mathcal{T}_h}+\tau_n\langle\widehat{\psi}_h,\theta\rangle_{\partial\mathcal{T}_h} &= 0,\end{aligned} \tag{A.3}$$

for any $(\boldsymbol{\kappa},\boldsymbol{\eta},\boldsymbol{\xi},\zeta,\boldsymbol{\mu},\theta) \in \boldsymbol{W}_h \times \boldsymbol{W}_h \times \boldsymbol{W}_h \times W_h \times \boldsymbol{M}_h(\mathbf{0}) \times M_h$. *Furthermore, if the stabilization parameters are positive on* $\partial\mathcal{T}_h$, *then the HDG solution* $(\mathbf{V}_h,\mathbf{E}_h,\mathbf{J}_h,U_h,\widehat{\mathbf{E}}_h,\widehat{U}_h)$ *exists and is uniquely defined.*

Proof. Since the square system above is linear and finite dimensional, it is sufficient to show that the trivial solution is the unique solution of (3.11) if $\mathbf{E}_D = \mathbf{V}_N = 0$. If we take $\boldsymbol{\kappa} = \mathbf{V}_h$, $\boldsymbol{\eta} = \mathbf{J}_h$, $\boldsymbol{\xi} = \mathbf{E}_h$, $\zeta = U_h$, $\boldsymbol{\mu} = \widehat{\mathbf{E}}_h$ and $\theta = \widehat{U}_h$ in (3.11), multiply the second equation by $-1/\omega_p^2$, the fourth and sixth by β^2/ω_p^2 and add them together, we arrive to

$$(\mathbf{V}_h,\mathbf{V}_h)_{\mathcal{T}_h} + \tau_t\langle(\mathbf{E}_h-\widehat{\mathbf{E}}_h)\times\mathbf{n},(\mathbf{E}_h-\widehat{\mathbf{E}}_h)\times\mathbf{n}\rangle_{\partial\mathcal{T}_h} + \frac{\beta^2}{\omega_p^2}(U_h,U_h)_{\mathcal{T}_h} +$$
$$\tau_n\langle U_h-\widehat{U}_h,U_h-\widehat{U}_h\rangle_{\partial\mathcal{T}_h} = \omega^2\varepsilon_\infty(\mathbf{E}_h,\mathbf{E}_h)_{\mathcal{T}_h} + \frac{\omega(\omega+i\gamma)}{\omega_p^2}(\mathbf{J}_h,\mathbf{J}_h)_{\mathcal{T}_h}.$$

Similarly, for the eigenproblem in (A.3) we have

$$(\mathbf{N}_h,\mathbf{N}_h)_{\mathcal{T}_h} + \tau_t\langle(\mathbf{Q}_h-\widehat{\mathbf{Q}}_h)\times\mathbf{n},(\mathbf{Q}_h-\widehat{\mathbf{Q}}_h)\times\mathbf{n}\rangle_{\partial\mathcal{T}_h} + \frac{\beta^2}{\omega_p^2}(\psi_h,\psi_h)_{\mathcal{T}_h} +$$
$$\tau_n\langle\psi_h-\widehat{\psi}_h,\psi_h-\widehat{\psi}_h\rangle_{\partial\mathcal{T}_h} = \lambda_1(\mathbf{Q}_h,\mathbf{Q}_h)_{\mathcal{T}_h} + \frac{\lambda_2}{\omega_p^2}(\mathbf{S}_h,\mathbf{S}_h)_{\mathcal{T}_h}.$$

It follows from the previous two equations that both $\mathbf{E}_h$ and $\mathbf{J}_h$ are zero; otherwise, then $\omega^2\varepsilon_\infty$ ans $\omega(\omega+i\gamma)$ must be eigenvalues of (A.3), which contradicts the hypothesis. As a consequence, we get

$$(\mathbf{V}_h,\mathbf{V}_h)_{\mathcal{T}_h} + \tau_t\langle\widehat{\mathbf{E}}_h\times\mathbf{n},\widehat{\mathbf{E}}_h\times\mathbf{n}\rangle_{\partial\mathcal{T}_h} + \frac{\beta^2}{\omega_p^2}(U_h,U_h)_{\mathcal{T}_h} + \tau_n\langle U_h-\widehat{U}_h,U_h-\widehat{U}_h\rangle_{\partial\mathcal{T}_h} = 0,$$

hence $\mathbf{V}_h = 0$, $\widehat{\mathbf{E}}_h = 0$, $U_h = 0$ and $\widehat{U}_h = 0$ since the stabilization constants are strictly positive. In consequence, the trivial solution is the unique solution of the HDG discretization, thus completing the proof. □

Appendix B

Deformation mapping for annular gap structure

In this appendix, we introduce and derive the deformation mapping that is employed throughout this dissertation for the periodic annular gap structure. Following the notation in [201], we assume a one-to-one mapping given by a diffeomorphism $\mathfrak{G}$ from a reference domain $\Omega_{\mathbf{r}} \in \mathbb{R}^3$ with coordinates $\mathbf{x}_{\mathbf{r}} = (x_{\mathbf{r}}, y_{\mathbf{r}}, z_{\mathbf{r}})$ to the physical domain $\Omega \in \mathbb{R}^3$ with coordinates $\mathbf{x} = (x, y, z)$. The mapping can be expressed as $\mathbf{x} = \mathfrak{G}(\mathbf{x}_{\mathbf{r}}, \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ parametrizes the mapping. The mapping deformation gradient and its Jacobian are defined $\mathcal{G} = \nabla_{\mathbf{r}}\mathfrak{G}$ and $g = \det \mathcal{G}$, respectively.

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 $\boldsymbol{\theta} = (\theta_1, \theta_2)$ that specify the modified radius $R + \theta_1$ and gap width θ_2. Fig. B-1 (left) shows the computational domain for the annular gap structure. To derive the mapping, we resort to polar coordinates for the reference domain $(\rho_{\mathbf{r}}, \alpha_{\mathbf{r}})$ and the physical domain (ρ, α). Furthermore, we have $\alpha_{\mathbf{r}} = \alpha$ since modifications occur only in the radial direction.

The physical coordinates relate to the reference coordinates as

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

In order to obtain $\rho(\mathbf{x}_{\mathbf{r}})$ we will first compute $\rho(\rho_{\mathbf{r}}, z_{\mathbf{r}})$ and then apply the chain rule to

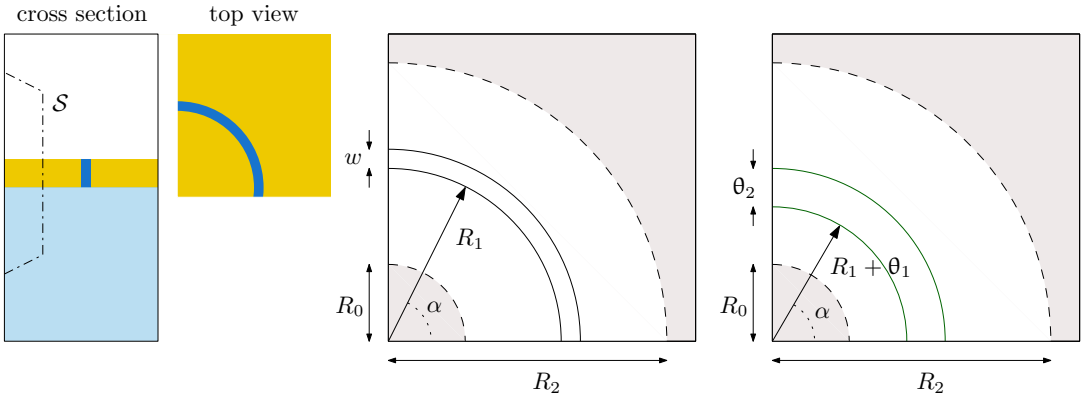


Figure B-1: Left: Computational domain with step function. Middle: reference domain. Right: physical domain.

compute the required derivatives. We propose to determine ρ using $\mathcal{C}^2$ cubic splines [54] in the radial direction. For both reference and physical domains in Fig. B-1 (middle and right), 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 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. In order to circumvent this limitation, we propose to use a linear interpolation between the spline and identity function using the continuous step function $\mathcal{S}$ shown in Fig B-1 (left) in dash-dot. This step function is one near the scatterer and continuosly fades to zero at a certain distance –enough to ensure an accurate representation of the deformation. All in all, the expression

that relates $\rho_{\mathbf{r}}$ to ρ reads

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

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

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

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

Bibliography

[1] *COMSOL Multiphysics Modeling Software*, http://www.comsol.com/.

[2] *Lumerical Solutions, Inc.*, http://www.lumerical.com/tcad-products/fdtd/.

[3] M. Ainsworth and J. Coyle, *Hierarchic hp-edge element families for Maxwell's equations on hybrid quadrilateral/triangular meshes*, Computer Methods in Applied Mechanics and Engineering, 190 (2001), pp. 6709–6733.

[4] ——, *Hierarchic finite element bases on unstructured tetrahedral meshes*, International Journal for Numerical Methods in Engineering, 58 (2003), pp. 2103–2130.

[5] M. N. Albunni, V. Rischmuller, T. Fritzsche, and B. Lohmann, *Model-order reduction of moving nonlinear electromagnetic devices*, IEEE Transactions on Magnetics, 44 (2008), pp. 1822–1829.

[6] A. Alonso and A. Valli, *An optimal domain decomposition preconditioner for low-frequency time-harmonic Maxwell equations*, Mathematics of Computation of the American Mathematical Society, 68 (1999), pp. 607–631.

[7] J. N. Anker, W. P. Hall, O. Lyandres, N. C. Shah, J. Zhao, and R. P. Van Duyne, *Biosensing with plasmonic nanosensors*, Nature materials, 7 (2008), pp. 442–453.

[8] G. Archer, A. Saltelli, and I. Sobol, *Sensitivity measures, ANOVA-like techniques and the use of bootstrap*, Journal of Statistical Computation and Simulation, 58 (1997), pp. 99–120.

[9] C. Argyropoulos, P.-Y. Chen, G. D'Aguanno, N. Engheta, and A. Alu, *Boosting optical nonlinearities in ε-near-zero plasmonic channels*, Physical Review B, 85 (2012), p. 045129.

[10] D. N. Arnold, F. Brezzi, B. Cockburn, and L. D. Marini, *Unified analysis of discontinuous Galerkin methods for elliptic problems*, SIAM Journal on Numerical Analysis, 39 (2002), pp. 1749–1779.

[11] N. W. Ashcroft and N. D. Mermin, *Solid State Physics (Holt, Rinehart and Winston, New York, 1976)*, Google Scholar, (2005), p. 403.

[12] H. A. Atwater and A. Polman, *Plasmonics for improved photovoltaic devices*, Nature materials, 9 (2010), pp. 205–213.

[13] I. Babuška and M. Zlámal, *Nonconforming elements in the finite element method with penalty*, SIAM Journal on Numerical Analysis, 10 (1973), pp. 863–875.

[14] F. Baida and D. Van Labeke, *Light transmission by subwavelength annular aperture arrays in metallic films*, Optics communications, 209 (2002), pp. 17–22.

[15] M. Barrault, Y. Maday, N. C. Nguyen, and A. T. Patera, *An 'empirical interpolation' method: application to efficient reduced-basis discretization of partial differential equations*, Comptes Rendus Mathematique, 339 (2004), pp. 667–672.

[16] F. Bassi and S. Rebay, *A high-order accurate discontinuous finite element method for the numerical solution of the compressible Navier–Stokes equations*, Journal of computational physics, 131 (1997), pp. 267–279.

[17] J.-P. Berenger, *A perfectly matched layer for the absorption of electromagnetic waves*, Journal of computational physics, 114 (1994), pp. 185–200.

[18] D. J. Bergman and M. I. Stockman, *Surface plasmon amplification by stimulated emission of radiation: quantum generation of coherent surface plasmons in nanosystems*, Physical review letters, 90 (2003), p. 027402.

[19] G. Berkooz, P. Holmes, and J. L. Lumley, *The proper orthogonal decomposition in the analysis of turbulent flows*, Annual review of fluid mechanics, 25 (1993), pp. 539–575.

[20] A. D. Boardman, *Electromagnetic surface modes*, John Wiley & Sons, 1982.

[21] C. F. Bohren and D. R. Huffman, *Absorption and scattering of light by small particles*, John Wiley & Sons, 2008.

[22] R. Boidin, T. Halenkovič, V. Nazabal, L. Beneš, and P. Němec, *Pulsed laser deposited alumina thin films*, Ceramics International, 42 (2016), pp. 1177–1182.

[23] A. Bossavit, *A rationale for 'edge-elements' in 3d fields computations*, IEEE Transactions on Magnetics, 24 (1988), pp. 74–79.

[24] A. Bossavit, *Solving Maxwell equations in a closed cavity, and the question of 'spurious modes'*, IEEE Transactions on magnetics, 26 (1990), pp. 702–705.

[25] S. Boyaval, C. Le Bris, T. Lelièvre, Y. Maday, N. C. Nguyen, and A. T. Patera, *Reduced basis techniques for stochastic problems*, Archives of Computational methods in Engineering, 17 (2010), pp. 435–454.

[26] S. I. Bozhevolnyi, V. S. Volkov, E. Devaux, J.-Y. Laluet, and T. W. Ebbesen, *Channel plasmon subwavelength waveguide components including interferometers and ring resonators*, Nature, 440 (2006), pp. 508–511.

[27] A. G. Brolo, R. Gordon, B. Leathem, and K. L. Kavanagh, *Surface plasmon sensor based on the enhanced light transmission through arrays of nanoholes in gold films*, Langmuir, 20 (2004), pp. 4813–4815.

[28] M. L. Brongersma, *Plasmonic photodetectors, photovoltaics, and hot-electron devices*, Proceedings of the IEEE, 104 (2016), pp. 2349–2361.

[29] T. Bui-Thanh, K. Willcox, and O. Ghattas, *Model reduction for large-scale systems with high-dimensional parametric input space*, SIAM Journal on Scientific Computing, 30 (2008), pp. 3270–3288.

[30] K. Busch, M. König, and J. Niegemann, *Discontinuous Galerkin methods in nanophotonics*, Laser & Photonics Reviews, 5 (2011), pp. 773–809.

[31] R. E. Caflisch, *Monte Carlo and quasi-Monte Carlo methods*, Acta numerica, 7 (1998), pp. 1–49.

[32] W. Cai and V. M. Shalaev, *Optical metamaterials*, vol. 10, Springer, 2010.

[33] J. P. Camden, J. A. Dieringer, J. Zhao, and R. P. Van Duyne, *Controlled plasmonic nanostructures for surface-enhanced spectroscopy and sensing*, Accounts of chemical research, 41 (2008), pp. 1653–1661.

[34] K. Catchpole and A. Polman, *Plasmonic solar cells*, Optics express, 16 (2008), pp. 21793–21800.

[35] W. Challener, C. Peng, A. Itagi, D. Karns, W. Peng, Y. Peng, X. Yang, X. Zhu, N. Gokemeijer, Y.-T. Hsia, et al., *Heat-assisted magnetic recording by a near-field transducer with efficient optical energy transfer*, Nature photonics, 3 (2009), pp. 220–224.

[36] D. Chandler-Horowitz and P. M. Amirtharaj, *High-accuracy, midinfrared ($450cm^{-1} < \omega < 4000cm^{-1}$) refractive index values of silicon*, Journal of Applied physics, 97 (2005), p. 123526.

[37] S. Chaturantabut and D. C. Sorensen, *Nonlinear model reduction via discrete empirical interpolation*, SIAM Journal on Scientific Computing, 32 (2010), pp. 2737–2764.

[38] P. Chen, A. Quarteroni, and G. Rozza, *Comparison between reduced basis and stochastic collocation methods for elliptic problems*, Journal of Scientific Computing, 59 (2014), pp. 187–216.

[39] X. Chen, H.-R. Park, N. C. Lindquist, J. Shaver, M. Pelton, and S.-H. Oh, *Squeezing millimeter waves through a single, nanometer-wide, centimeter-long slit*, Scientific reports, 4 (2014), p. 6722.

[40] X. Chen, H.-R. Park, M. Pelton, X. Piao, N. C. Lindquist, H. Im, Y. J. Kim, J. S. Ahn, K. J. Ahn, N. Park, et al., *Atomic layer lithography of wafer-scale nanogap arrays for extreme confinement of electromagnetic waves*, Nature communications, 4 (2013).

[41] Y. Chen, J. S. Hesthaven, and 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.

[42] Y. Chen, J. S. Hesthaven, Y. Maday, and J. Rodríguez, *Improved successive constraint method based a posteriori error estimate for reduced basis approximation of 2d Maxwell's problem*, ESAIM: Mathematical Modelling and Numerical Analysis, 43 (2009), pp. 1099–1116.

[43] ——, *Certified reduced basis methods and output bounds for the harmonic Maxwell's equations*, SIAM Journal on Scientific Computing, 32 (2010), pp. 970–996.

[44] C. Ciracì and F. Della Sala, *Quantum hydrodynamic theory for plasmonics: Impact of the electron density tail*, Physical Review B, 93 (2016), p. 205405.

[45] C. Ciracì, R. Hill, J. Mock, Y. Urzhumov, A. Fernández-Domínguez, S. Maier, J. Pendry, A. Chilkoti, and D. Smith, *Probing the ultimate limits of plasmonic enhancement*, Science, 337 (2012), pp. 1072–1074.

[46] C. Ciracì, J. B. Pendry, and D. R. Smith, *Hydrodynamic model for plasmonics: a macroscopic approach to a microscopic problem*, ChemPhysChem, 14 (2013), pp. 1109–1116.

[47] B. Cockburn, B. Dong, and J. Guzmán, *A superconvergent LDG-hybridizable Galerkin method for second-order elliptic problems*, Mathematics of Computation, 77 (2008), pp. 1887–1916.

[48] B. Cockburn, J. Gopalakrishnan, and R. Lazarov, *Unified hybridization of discontinuous Galerkin, mixed, and continuous Galerkin methods for second order elliptic problems*, SIAM Journal on Numerical Analysis, 47 (2009), pp. 1319–1365.

[49] B. Cockburn, J. Gopalakrishnan, N. C. Nguyen, J. Peraire, and F.-J. Sayas, *Analysis of HDG methods for Stokes flow*, Mathematics of Computation, 80 (2011), pp. 723–760.

[50] B. Cockburn, J. Gopalakrishnan, and F.-J. Sayas, *A projection-based error analysis of HDG methods*, Mathematics of Computation, 79 (2010), pp. 1351–1367.

[51] B. Cockburn, J. Guzmán, S.-C. Soon, and H. K. Stolarski, *An analysis of the embedded discontinuous Galerkin method for second-order elliptic problems*, SIAM Journal on Numerical Analysis, 47 (2009), pp. 2686–2707.

[52] B. Cockburn, J. Guzmán, and H. Wang, *Superconvergent discontinuous Galerkin methods for second-order elliptic problems*, Mathematics of Computation, 78 (2009), pp. 1–24.

[53] B. Cockburn and C.-W. Shu, *The local discontinuous Galerkin method for time-dependent convection-diffusion systems*, SIAM Journal on Numerical Analysis, 35 (1998), pp. 2440–2463.

[54] C. De Boor, C. De Boor, E.-U. Mathématicien, C. De Boor, and C. De Boor, *A practical guide to splines*, vol. 27, Springer-Verlag New York, 1978.

[55] V. De La Rubia, U. Razafison, and Y. Maday, *Reliable fast frequency sweep for microwave devices via the reduced-basis method*, IEEE Transactions on Microwave Theory and Techniques, 57 (2009), pp. 2923–2937.

[56] L. Demkowicz, *Computing with hp-adaptive finite elements: Volume 1 one and two dimensional elliptic and maxwell problems*, CRC Press, 2006.

[57] L. Demkowicz and L. Vardapetyan, *Modeling of electromagnetic absorption/scattering problems using hp-adaptive finite elements*, Computer Methods in Applied Mechanics and Engineering, 152 (1998), pp. 103–124.

[58] E. B. Dickerson, E. C. Dreaden, X. Huang, I. H. El-Sayed, H. Chu, S. Pushpanketh, J. F. McDonald, and M. A. El-Sayed, *Gold nanorod assisted near-infrared plasmonic photothermal therapy (PPTT) of squamous cell carcinoma in mice*, Cancer letters, 269 (2008), pp. 57–66.

[59] M. Dodge, *Refractive index*, Handbook of Laser Science and Technology, 4 (1986), pp. 21–47.

[60] M. Drohmann and K. Carlberg, *The ROMES method for statistical modeling of reduced-order-model error*, SIAM/ASA Journal on Uncertainty Quantification, 3 (2015), pp. 116–145.

[61] P. Drude, *Zur elektronentheorie der metalle*, Annalen der Physik, 306 (1900), pp. 566–613.

[62] L. Dykman and N. Khlebtsov, *Gold nanoparticles in biomedical applications: recent advances and perspectives*, Chemical Society Reviews, 41 (2012), pp. 2256–2282.

[63] T. W. Ebbesen, H. J. Lezec, H. Ghaemi, T. Thio, and P. Wolff, *Extraordinary optical transmission through sub-wavelength hole arrays*, Nature, 391 (1998), pp. 667–669.

[64] J. L. Eftang and A. T. Patera, *Port reduction in parametrized component static condensation: approximation and a posteriori error estimation*, International Journal for Numerical Methods in Engineering, 96 (2013), pp. 269–302.

[65] A. Eguiluz, S. Ying, and J. Quinn, *Influence of the electron density profile on surface plasmons in a hydrodynamic model*, Physical Review B, 11 (1975), p. 2118.

[66] I. H. El-Sayed, X. Huang, and M. A. El-Sayed, *Surface plasmon resonance scattering and absorption of anti-EGFR antibody conjugated gold nanoparticles in cancer diagnostics: applications in oral cancer*, Nano letters, 5 (2005), pp. 829–834.

[67] B. Engquist and A. Majda, *Absorbing boundary conditions for numerical simulation of waves*, Proceedings of the National Academy of Sciences, 74 (1977), pp. 1765–1766.

[68] B. Epureanu, E. Dowell, and K. Hall, *A parametric analysis of reduced order models of potential flows in turbomachinery using proper orthogonal decomposition*, in Proceedings of ASME turbo expo, vol. 2001, 2001.

[69] R. Esteban, A. G. Borisov, P. Nordlander, and J. Aizpurua, *Bridging quantum and classical plasmonics with a quantum-corrected model*, Nature communications, 3 (2012), p. 825.

[70] R. Esteban, A. Zugarramurdi, P. Zhang, P. Nordlander, F. J. García-Vidal, A. G. Borisov, and J. Aizpurua, *A classical treatment of optical tunneling in plasmonic gaps: extending the quantum corrected model to practical situations*, Faraday discussions, 178 (2015), pp. 151–183.

[71] W. Fan, S. Zhang, B. Minhas, K. J. Malloy, and S. Brueck, *Enhanced infrared transmission through subwavelength coaxial metallic arrays*, Physical review letters, 94 (2005), p. 033902.

[72] B. Feeny and R. Kappagantu, *On the physical interpretation of proper orthogonal modes in vibrations*, Journal of sound and vibration, 211 (1998), pp. 607–616.

[73] P. Feldmann and R. W. Freund, *Efficient linear circuit analysis by Padé approximation via the Lanczos process*, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 14 (1995), pp. 639–649.

[74] P. Fernández, N. C. Nguyen, X. Roca, and J. Peraire, *Implicit large-eddy simulation of compressible flows using the Interior Embedded Discontinuous Galerkin method*, in 54th AIAA Aerospace Sciences Meeting, 2016, p. 1332.

[75] A. Fernández-Domínguez, A. Wiener, F. García-Vidal, S. Maier, and J. Pendry, *Transformation-optics description of nonlocal effects in plasmonic nanostructures*, Physical review letters, 108 (2012), p. 106802.

[76] V. E. Ferry, L. A. Sweatlock, D. Pacifici, and H. A. Atwater, *Plasmonic nanostructure design for efficient light coupling into solar cells*, Nano letters, 8 (2008), pp. 4391–4397.

[77] V. E. Ferry, M. A. Verschuuren, H. B. Li, E. Verhagen, R. J. Walters, R. E. Schropp, H. A. Atwater, and A. Polman, *Light trapping in ultrathin plasmonic solar cells*, Optics express, 18 (2010), pp. A237–A245.

[78] J. Fink and W. Rheinboldt, *On the error behavior of the reduced basis technique for nonlinear finite element approximations*, ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik, 63 (1983), pp. 21–28.

[79] J. M. Fitzgerald, P. Narang, R. V. Craster, S. A. Maier, and V. Giannini, *Quantum plasmonics*, Proceedings of the IEEE, 104 (2016), pp. 2307–2322.

[80] H. G. Frey, S. Witt, K. Felderer, and R. Guckenberger, *High-resolution imaging of single fluorescent molecules with the optical near-field of a metal tip*, Physical review letters, 93 (2004), p. 200801.

[81] F. J. Garcia-Vidal, L. Martin-Moreno, T. Ebbesen, and L. Kuipers, *Light passing through subwavelength apertures*, Reviews of Modern Physics, 82 (2010), p. 729.

[82] W. C. Gibson, *The method of moments in electromagnetics*, vol. 1, Chapman & Hall/CRC London, UK, 2008.

[83] M. B. Giles, *Multilevel Monte Carlo path simulation*, Operations Research, 56 (2008), pp. 607–617.

[84] D. Givoli, *High-order local non-reflecting boundary conditions: a review*, Wave motion, 39 (2004), pp. 319–326.

[85] C. Greif and D. Schötzau, *Preconditioners for the discretized time-harmonic maxwell equations in mixed form*, Numerical Linear Algebra with Applications, 14 (2007), pp. 281–297.

[86] M. A. Grepl, *Reduced-basis approximations for time-dependent partial differential equations: application to optimal control*, PhD thesis, Massachusetts Institute of Technology, 2005.

[87] M. A. Grepl, Y. Maday, N. C. Nguyen, and A. T. Patera, *Efficient reduced-basis treatment of nonaffine and nonlinear partial differential equations*, ESAIM: Mathematical Modelling and Numerical Analysis, 41 (2007), pp. 575–605.

[88] M. A. Grepl and A. T. Patera, *A posteriori error bounds for reduced-basis approximations of parametrized parabolic partial differential equations*, ESAIM: Mathematical Modelling and Numerical Analysis, 39 (2005), pp. 157–181.

[89] A. Grigorenko, M. Polini, and K. Novoselov, *Graphene plasmonics*, Nature photonics, 6 (2012), pp. 749–758.

[90] E. J. Grimme, *Krylov projection methods for model reduction*, PhD thesis, Citeseer, 1997.

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

[92] M. Groner, J. Elam, F. Fabreguette, and 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 (2002), pp. 186–197.

[93] S. Gugercin and A. C. Antoulas, *A survey of model reduction by balanced truncation and some new results*, International Journal of Control, 77 (2004), pp. 748–766.

[94] B. Haasdonk, K. Urban, and B. Wieland, *Reduced basis methods for parameterized partial differential equations with stochastic influences using the Karhunen–Loève expansion*, SIAM/ASA Journal on Uncertainty Quantification, 1 (2013), pp. 79–105.

[95] T. Hagstrom, *Radiation boundary conditions for the numerical simulation of waves*, Acta numerica, 8 (1999), pp. 47–106.

[96] N. J. Halas, S. Lal, W.-S. Chang, S. Link, and P. Nordlander, *Plasmons in strongly coupled metallic nanostructures*, Chemical reviews, 111 (2011), pp. 3913–3961.

[97] K. C. Hall, J. P. Thomas, and E. H. Dowell, *Proper orthogonal decomposition technique for transonic unsteady aerodynamic flows*, AIAA journal, 38 (2000), pp. 1853–1862.

[98] R. F. Harrington and J. L. Harrington, *Field computation by moment methods*, Oxford University Press, 1996.

[99] A. Hartschuh, E. J. Sánchez, X. S. Xie, and L. Novotny, *High-resolution near-field Raman microscopy of single-walled carbon nanotubes*, Physical Review Letters, 90 (2003), p. 095503.

[100] M. W. Hess and P. Benner, *Fast evaluation of time–harmonic Maxwell's equations using the reduced basis method*, IEEE Transactions on Microwave Theory and Techniques, 61 (2013), pp. 2265–2274.

[101] J. S. Hesthaven and T. Warburton, *Nodal high-order methods on unstructured grids: I. Time-domain solution of Maxwell's equations*, Journal of Computational Physics, 181 (2002), pp. 186–221.

[102] K. R. Hiremath, L. Zschiedrich, and F. Schmidt, *Numerical solution of nonlocal hydrodynamic Drude model for arbitrary shaped nano-plasmonic structures using Nédélec finite elements*, Journal of Computational Physics, 231 (2012), pp. 5890–5896.

[103] P. Holmes, J. L. Lumley, and G. Berkooz, *Turbulence, coherent structures, dynamical systems and symmetry*, Cambridge university press, 1998.

[104] X. Huang, I. H. El-Sayed, W. Qian, and M. A. El-Sayed, *Cancer cell imaging and photothermal therapy in the near-infrared region by using gold nanorods*, Journal of the American Chemical Society, 128 (2006), pp. 2115–2120.

[105] X. Huang, P. K. Jain, I. H. El-Sayed, and M. A. El-Sayed, *Plasmonic photothermal therapy (PPTT) using gold nanoparticles*, Lasers in medical science, 23 (2008), p. 217.

[106] X. Huang, S. Neretina, and M. A. El-Sayed, *Gold nanorods: from synthesis and properties to biological and biomedical applications*, Advanced Materials, 21 (2009), pp. 4880–4910.

[107] A. J. Huber, F. Keilmann, J. Wittborn, J. Aizpurua, and R. Hillenbrand, *Terahertz near-field nanoscopy of mobile carriers in single semiconductor nanodevices*, Nano letters, 8 (2008), pp. 3766–3770.

[108] A. Huerta, A. Angeloski, X. Roca, and J. Peraire, *Efficiency of high-order elements for continuous and discontinuous Galerkin methods*, International Journal for numerical methods in Engineering, 96 (2013), pp. 529–560.

[109] E. S. Hung and S. D. Senturia, *Generating efficient dynamical models for microelectromechanical systems from a few finite-element simulation runs*, Journal of Microelectromechanical Systems, 8 (1999), pp. 280–289.

[110] D. B. P. Huynh, D. J. Knezevic, Y. Chen, J. S. Hesthaven, and A. T. Patera, *A natural-norm successive constraint method for inf-sup lower bounds*, Computer Methods in Applied Mechanics and Engineering, 199 (2010), pp. 1963–1975.

[111] D. B. P. Huynh, D. J. Knezevic, and A. T. Patera, *A static condensation reduced basis element method: approximation and a posteriori error estimation*, ESAIM: Mathematical Modelling and Numerical Analysis, 47 (2013), pp. 213–251.

[112] ——, *A static condensation reduced basis element method: Complex problems*, Computer Methods in Applied Mechanics and Engineering, 259 (2013), pp. 197–216.

[113] D. B. P. Huynh, G. Rozza, S. Sen, and A. T. Patera, *A successive constraint linear optimization method for lower bounds of parametric coercivity and inf–sup stability constants*, Comptes Rendus Mathematique, 345 (2007), pp. 473–478.

[114] H. Im, K. C. Bantz, N. C. Lindquist, C. L. Haynes, and S.-H. Oh, *Vertically oriented sub-10-nm plasmonic nanogap arrays*, Nano letters, 10 (2010), pp. 2231–2236.

[115] M. Jablan, H. Buljan, and M. Soljačić, *Plasmonics in graphene at infrared frequencies*, Physical review B, 80 (2009), p. 245435.

[116] P. K. Jain, S. Eustis, and M. A. El-Sayed, *Plasmon coupling in nanorod assemblies: optical absorption, discrete dipole approximation simulation, and exciton-coupling model*, The Journal of Physical Chemistry B, 110 (2006), pp. 18243–18253.

[117] X. Ji, W. Cai, and P. Zhang, *High-order DGTD methods for dispersive Maxwell's equations and modelling of silver nanowire coupling*, International journal for numerical methods in engineering, 69 (2007), pp. 308–325.

[118] J.-M. Jin, *The finite element method in electromagnetics*, John Wiley & Sons, 2014.

[119] P. B. Johnson and R.-W. Christy, *Optical constants of the noble metals*, Physical review B, 6 (1972), p. 4370.

[120] S. G. Johnson, *Notes on perfectly matched layers (PMLs)*, Lecture notes, Massachusetts Institute of Technology, Massachusetts, 5 (2008), p. 2.

[121] S. G. Johnson and J. D. Joannopoulos, *Block-iterative frequency-domain methods for Maxwell's equations in a planewave basis*, Optics express, 8 (2001), pp. 173–190.

[122] I. Jolliffe, *Principal component analysis*, Wiley Online Library, 2002.

[123] L. Ju, B. Geng, J. Horng, C. Girit, M. Martin, Z. Hao, H. A. Bechtel, X. Liang, A. Zettl, Y. R. Shen, et al., *Graphene plasmonics for tunable terahertz metamaterials*, Nature nanotechnology, 6 (2011), pp. 630–634.

[124] A. Kabashin, P. Evans, S. Pastkovsky, W. Hendren, G. Wurtz, R. Atkinson, R. Pollard, V. Podolskiy, and A. Zayats, *Plasmonic nanorod metamaterials for biosensing*, Nature materials, 8 (2009), pp. 867–871.

[125] T. Kalkbrenner, M. Ramstein, J. Mlynek, and V. Sandoghdar, *A single gold particle as a probe for apertureless scanning near-field optical microscopy*, Journal of Microscopy, 202 (2001), pp. 72–76.

[126] S. Kawata, Y. Inouye, and P. Verma, *Plasmonics for near-field nano-imaging and superlensing*, Nature Photonics, 3 (2009), pp. 388–394.

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

[128] J. Kischkat, S. Peters, B. Gruska, M. Semtsiv, M. Chashnikova, M. Klinkmüller, O. Fedosenko, S. Machulik, A. Aleksandrova, G. Monastyrskyi, et al., *Mid-infrared optical properties of thin films of aluminum oxide, titanium dioxide, silicon dioxide, aluminum nitride, and silicon nitride*, Applied optics, 51 (2012), pp. 6789–6798.

[129] F. H. Koppens, D. E. Chang, and F. J. García de Abajo, *Graphene plasmonics: a platform for strong light–matter interactions*, Nano letters, 11 (2011), pp. 3370–3377.

[130] J. P. Kottmann and O. J. Martin, *Plasmon resonant coupling in metallic nanowires*, Optics Express, 8 (2001), pp. 655–663.

[131] M. Kowalski and J.-M. Jin, *Karhunen-Loève based model order reduction of nonlinear systems*, in Antennas and Propagation Society International Symposium, 2002. IEEE, vol. 2, IEEE, 2002, pp. 552–555.

[132] K. S. Kunz and R. J. Luebbers, *The finite difference time domain method for electromagnetics*, CRC press, 1993.

[133] S. Lal, S. Link, and N. J. Halas, *Nano-optics from sensing to waveguiding*, Nature photonics, 1 (2007), pp. 641–648.

[134] S. Lanteri and C. Scheid, *Convergence of a discontinuous Galerkin scheme for the mixed time-domain Maxwell's equations in dispersive media*, IMA Journal of Numerical Analysis, 33 (2013), pp. 432–459.

[135] H. J. Lezec, A. Degiron, E. Devaux, R. Linke, L. Martin-Moreno, F. Garcia-Vidal, and T. Ebbesen, *Beaming light from a subwavelength aperture*, Science, 297 (2002), pp. 820–822.

[136] J.-H. Li, M. Hayashi, and G.-Y. Guo, *Plasmonic excitations in quantum-sized sodium nanoparticles studied by time-dependent density functional calculations*, Physical Review B, 88 (2013), p. 155437.

[137] L. Li, Y. Yu, G. J. Ye, Q. Ge, X. Ou, H. Wu, D. Feng, X. H. Chen, and Y. Zhang, *Black phosphorus field-effect transistors*, Nature nanotechnology, 9 (2014), pp. 372–377.

[138] Y. Liang, H. Lee, S. Lim, W. Lin, K. Lee, and C. Wu, *Proper orthogonal decomposition and its applications–Part I: Theory*, Journal of Sound and vibration, 252 (2002), pp. 527–544.

[139] Y. Liang, W. Lin, H. Lee, S. Lim, K. Lee, and H. Sun, *Proper orthogonal decomposition and its applications–Part II: Model reduction for MEMS dynamical analysis*, Journal of Sound and Vibration, 256 (2002), pp. 515–532.

[140] J. Lin, S.-H. Oh, H.-M. Nguyen, and F. Reitich, *Field enhancement and saturation of millimeter waves inside a metallic nanogap*, Optics express, 22 (2014), pp. 14402–14410.

[141] N. Liu, M. Mesch, T. Weiss, M. Hentschel, and H. Giessen, *Infrared perfect absorber and its application as plasmonic sensor*, Nano letters, 10 (2010), pp. 2342–2348.

[142] N. Liu, M. L. Tang, M. Hentschel, H. Giessen, and A. P. Alivisatos, *Nanoantenna-enhanced gas sensing in a single tailored nanofocus*, Nature materials, 10 (2011), pp. 631–636.

[143] W. Loh, *On latin hypercube sampling*, Ann. Statist., 24 (1996), pp. 2058–2080.

[144] T. Low, R. Roldán, H. Wang, F. Xia, P. Avouris, L. M. Moreno, and F. Guinea, *Plasmons and screening in monolayer and multilayer black phosphorus*, Physical review letters, 113 (2014), p. 106802.

[145] T. Lu, P. Zhang, and W. Cai, *Discontinuous Galerkin methods for dispersive and lossy Maxwell's equations and PML boundary conditions*, Journal of Computational Physics, 200 (2004), pp. 549–580.

[146] J. L. Lumley, *Stochastic tools in turbulence*, Courier Corporation, 2007.

[147] H. V. Ly and H. T. Tran, *Modeling and control of physical processes using proper orthogonal decomposition*, Mathematical and computer modelling, 33 (2001), pp. 223–236.

[148] L. Machiels, Y. Maday, I. B. Oliveira, A. T. Patera, and D. V. Rovas, *Output bounds for reduced-basis approximations of symmetric positive definite eigenvalue problems*, Comptes Rendus de l'Academie des Sciences Series I Mathematics, 331 (2000), pp. 153–158.

[149] L. Machiels, Y. Maday, and A. T. Patera, *Output bounds for reduced-order approximations of elliptic partial differential equations*, Computer methods in applied mechanics and engineering, 190 (2001), pp. 3413–3426.

[150] Y. Maday, A. T. Patera, and D. V. Rovas, *A blackbox reduced-basis output bound method for noncoercive linear problems*, Studies in Mathematics and its applications, 31 (2002), pp. 533–569.

[151] Y. Maday, A. T. Patera, and G. Turinici, *Global a priori convergence theory for reduced-basis approximations of single-parameter symmetric coercive elliptic partial differential equations*, Comptes Rendus Mathematique, 335 (2002), pp. 289–294.

[152] Y. Maday and E. M. Ronquist, *The reduced basis element method: application to a thermal fin problem*, SIAM Journal on Scientific Computing, 26 (2004), pp. 240–258.

[153] S. A. Maier, *Plasmonics: fundamentals and applications*, Springer Science & Business Media, 2007.

[154] S. A. Maier and H. A. Atwater, *Plasmonics: Localization and guiding of electromagnetic energy in metal/dielectric structures*, Journal of Applied Physics, 98 (2005), p. 10.

[155] S. A. Maier, M. L. Brongersma, P. G. Kik, S. Meltzer, A. A. Requicha, and H. A. Atwater, *Plasmonics–a route to nanoscale optical devices*, Advanced Materials, 13 (2001), pp. 1501–1505.

[156] M. Mansuripur, A. Zakharian, A. Lesuffleur, S.-H. Oh, R. Jones, N. Lindquist, H. Im, A. Kobyakov, and J. Moloney, *Plasmonic nano-structures for optical data storage*, Optics express, 17 (2009), pp. 14001–14014.

[157] J. M. McMahon, S. K. Gray, and G. C. Schatz, *Calculating nonlocal optical properties of structures with arbitrary shape*, Physical Review B, 82 (2010), p. 035423.

[158] M. Meyer and H. G. Matthies, *Efficient model reduction in non-linear dynamics using the Karhunen-Loève expansion and dual-weighted-residual methods*, Computational Mechanics, 31 (2003), pp. 179–191.

[159] E. K. Miller, *A selective survey of computational electromagnetics*, IEEE Transactions on Antennas and Propagation, 36 (1988), pp. 1281–1305.

[160] P. Monk, *Finite element methods for Maxwell's equations*, Oxford University Press, 2003.

[161] B. Moore, *Principal component analysis in linear systems: Controllability, observability, and model reduction*, IEEE transactions on automatic control, 26 (1981), pp. 17–32.

[162] A. Moosavi, R. Stefanescu, and A. Sandu, *Efficient construction of local parametric reduced order models using machine learning techniques*, arXiv preprint arXiv:1511.02909, (2015).

[163] A. Moreau, C. Ciraci, and D. R. Smith, *Impact of nonlocal response on metallodielectric multilayers and optical patch antennas*, Physical Review B, 87 (2013), p. 045401.

[164] N. A. Mortensen, S. Raza, M. Wubs, T. Søndergaard, and S. I. Bozhevolnyi, *A generalized non-local optical response theory for plasmonic nanostructures*, Nature communications, 5 (2014).

[165] G. Mur, *Absorbing boundary conditions for the finite-difference approximation of the time-domain electromagnetic-field equations*, Electromagnetic Compatibility, IEEE Transactions on, (1981), pp. 377–382.

[166] M. Naftaly and R. Miles, *Terahertz time-domain spectroscopy: A new tool for the study of glasses in the far infrared*, Journal of non-crystalline solids, 351 (2005), pp. 3341–3346.

[167] J.-C. Nédélec, *Mixed finite elements in* $\mathbb{R}^3$, Numerische Mathematik, 35 (1980), pp. 315–341.

[168] ——, *A new family of mixed finite elements in* $\mathbb{R}^3$, Numerische Mathematik, 50 (1986), pp. 57–81.

[169] N. Nguyen, J. Peraire, and B. Cockburn, *A hybridizable discontinuous galerkin method for the incompressible navier-stokes equations*, in 48th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition, 2010, p. 362.

[170] N. C. Nguyen, *Reduced-basis approximations and a posteriori error bounds for non-affine and nonlinear partial differential equations: application to inverse analysis*, PhD thesis, Singapore-MIT Alliance, June 2005.

[171] N. C. Nguyen, *A multiscale reduced-basis method for parametrized elliptic partial differential equations with multiple scales*, Journal of Computational Physics, 227 (2008), pp. 9807–9822.

[172] N. C. Nguyen, A. T. Patera, and J. Peraire, *A 'best points' interpolation method for efficient approximation of parametrized functions*, International Journal for Numerical Methods in Engineering, 73 (2008), pp. 521–543.

[173] N. C. Nguyen and J. Peraire, *Hybridizable discontinuous Galerkin methods for partial differential equations in continuum mechanics*, Journal of Computational Physics, 231 (2012), pp. 5955–5988.

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

[175] ——, *An implicit high-order hybridizable discontinuous Galerkin method for nonlinear convection-diffusion equations*, J. Comput. Phys., 228 (2009), pp. 8841–8855.

[176] ——, *A hybridizable discontinuous Galerkin method for Stokes flow*, Computer Methods in Applied Mechanics and Engineering, 199 (2010), pp. 582–597.

[177] ——, *An implicit high-order hybridizable discontinuous Galerkin method for the incompressible Navier-Stokes equations*, J. Comp. Physics, 230 (2011), pp. 1147–1170.

[178] ——, *High-order implicit hybridizable discontinuous Galerkin methods for acoustics and elastodynamics*, J. Comp. Physics, 230 (2011), pp. 3695–3718.

[179] ——, *Hybridizable discontinuous Galerkin methods for the time-harmonic Maxwell's equations*, J. Comp. Physics, 230 (2011), pp. 7151–7175.

[180] ——, *A class of embedded discontinuous Galerkin methods for computational fluid dynamics*, Journal of Computational Physics, 302 (2015), pp. 674–692.

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

[182] H. Niederreiter, *Random number generation and quasi-Monte Carlo methods*, SIAM, Philadelphia, 1992.

[183] A. K. Noor and J. M. Peters, *Reduced basis technique for nonlinear analysis of structures*, AIAA Journal, 18 (1980), pp. 455–462.

[184] ——, *Multiple-parameter reduced basis technique for bifurcation and post-buckling analyses of composite plates*, International Journal for Numerical Methods in Engineering, 19 (1983), pp. 1783–1803.

[185] P. Nordlander, C. Oubre, E. Prodan, K. Li, and M. Stockman, *Plasmon hybridization in nanoparticle dimers*, Nano letters, 4 (2004), pp. 899–903.

[186] L. Novotny and N. Van Hulst, *Antennas for light*, Nature photonics, 5 (2011), pp. 83–90.

[187] D. O'Connor and A. V. Zayats, *Data storage: The third plasmonic revolution*, Nature nanotechnology, 5 (2010), pp. 482–483.

[188] J. F. O'Hara, R. Singh, I. Brener, E. Smirnova, J. Han, A. J. Taylor, and W. Zhang, *Thin-film sensing with planar terahertz metamaterials: sensitivity and limitations*, Optics Express, 16 (2008), pp. 1786–1795.

[189] J. F. O'Hara, W. Withayachumnankul, and I. Al-Naib, *A review on thin-film sensing with terahertz waves*, Journal of Infrared, Millimeter, and Terahertz Waves, 33 (2012), pp. 245–291.

[190] R. L. Olmon, B. Slovick, T. W. Johnson, D. Shelton, S.-H. Oh, G. D. Boreman, and M. B. Raschke, *Optical dielectric function of gold*, Physical Review B, 86 (2012), p. 235147.

[191] G. Onida, L. Reining, and A. Rubio, *Electronic excitations: density-functional versus many-body Green's-function approaches*, Reviews of Modern Physics, 74 (2002), p. 601.

[192] M. Ordal, L. Long, R. Bell, S. Bell, R. Bell, R. Alexander, and C. Ward, *Optical properties of the metals Al, Co, Cu, Au, Fe, Pb, Ni, Pd, Pt, Ag, Ti, and W in the infrared and far infrared*, Applied Optics, 22 (1983), pp. 1099–1119.

[193] M. A. Ordal, R. J. Bell, R. W. Alexander, L. L. Long, and 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.*, Applied optics, 24 (1985), pp. 4493–4499.

[194] A. F. Oskooi, D. Roundy, M. Ibanescu, P. Bermel, J. Joannopoulos, and S. G. Johnson, *MEEP: A flexible free-software package for electromagnetic simulations by the FDTD method*, Computer Physics Communications, 181 (2010), pp. 687–702.

[195] R. F. Oulton, V. J. Sorger, D. Genov, D. Pile, and X. Zhang, *A hybrid plasmonic waveguide for subwavelength confinement and long-range propagation*, Nature Photonics, 2 (2008), pp. 496–500.

[196] E. Ozbay, *Plasmonics: merging photonics and electronics at nanoscale dimensions*, science, 311 (2006), pp. 189–193.

[197] E. D. Palik, *Handbook of optical constants of solids*, vol. 3, Academic press, 1998.

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

[199] B. Peherstorfer, D. Butnaru, K. Willcox, and H.-J. Bungartz, *Localized discrete empirical interpolation method*, SIAM Journal on Scientific Computing, 36 (2014), pp. A168–A192.

[200] J. Peraire, N. C. Nguyen, and B. Cockburn, *An embedded discontinuous Galerkin method for the compressible Euler and Navier-Stokes equations*, in 20th AIAA Computational Fluid Dynamics Conference, 2011, p. 3228.

[201] P.-O. Persson, J. Bonet, and J. Peraire, *Discontinuous Galerkin solution of the Navier–Stokes equations on deformable domains*, Computer Methods in Applied Mechanics and Engineering, 198 (2009), pp. 1585–1595.

[202] D. F. Pile, T. Ogawa, D. K. Gramotnev, T. Okamoto, M. Haraguchi, M. Fukui, and S. Matsuo, *Theoretical and experimental investigation of strongly localized plasmons on triangular metal wedges for subwavelength waveguiding*, Applied Physics Letters, 87 (2005), p. 061106.

[203] J. Pitarke, V. Silkin, E. Chulkov, and P. Echenique, *Theory of surface plasmons and surface-plasmon polaritons*, Reports on progress in physics, 70 (2006), p. 1.

[204] T. Porsching, *Estimation of the error in the reduced basis method solution of nonlinear equations*, Mathematics of Computation, 45 (1985), pp. 487–496.

[205] Y. Poujet, J. Salvi, and F. I. Baida, *90% extraordinary optical transmission in the visible range through annular aperture metallic arrays*, Optics letters, 32 (2007), pp. 2942–2944.

[206] C. Prud'homme, D. V. Rovas, K. Veroy, L. Machiels, Y. Maday, A. T. Patera, and G. Turinici, *Reliable real-time solution of parametrized partial differential equations: Reduced-basis output bound methods*, Journal of Fluids Engineering, 124 (2002), pp. 70–80.

[207] D. Rabinovich, D. Givoli, and E. Bécache, *Comparison of high-order absorbing boundary conditions and perfectly matched layers in the frequency domain*, International Journal for Numerical Methods in Biomedical Engineering, 26 (2010), pp. 1351–1369.

[208] H. Raether, *Excitation of plasmons and interband transitions by electrons*, Springer, 1980.

[209] S. M. Rao, D. R. Wilton, and A. W. Glisson, *Electromagnetic scattering by surfaces of arbitrary shape*, Antennas and Propagation, IEEE Transactions on, 30 (1982), pp. 409–418.

[210] P.-A. Raviart and J.-M. Thomas, *A mixed finite element method for 2-nd order elliptic problems*, in Mathematical aspects of finite element methods, Springer, 1977, pp. 292–315.

[211] S. Ravindran, *A reduced-order approach for optimal control of fluids using proper orthogonal decomposition*, International journal for numerical methods in fluids, 34 (2000), pp. 425–448.

[212] S. Raza, S. I. Bozhevolnyi, M. Wubs, and N. A. Mortensen, *Nonlocal optical response in metallic nanostructures*, Journal of Physics: Condensed Matter, 27 (2015), p. 183204.

[213] S. Raza, N. Stenger, S. Kadkhodazadeh, S. V. Fischer, N. Kostesha, A.-P. Jauho, A. Burrows, M. Wubs, and N. A. Mortensen, *Blueshift of the surface plasmon resonance in silver nanoparticles studied with EELS*, Nanophotonics, 2 (2013), pp. 131–138.

[214] S. Raza, M. Wubs, S. I. Bozhevolnyi, and N. A. Mortensen, *Nonlocal study of ultimate plasmon hybridization*, Optics letters, 40 (2015), pp. 839–842.

[215] T. Reiners, C. Ellert, M. Schmidt, and H. Haberland, *Size dependence of the optical response of spherical sodium clusters*, Physical review letters, 74 (1995), p. 1558.

[216] M. Rewienski and J. White, *A trajectory piecewise-linear approach to model order reduction and fast simulation of nonlinear circuits and micromachined devices*, IEEE Transactions on computer-aided design of integrated circuits and systems, 22 (2003), pp. 155–170.

[217] X. Roca, N. C. Nguyen, and J. Peraire, *Scalable parallelization of the hybridized discontinuous Galerkin method for compressible flow*, in 21st AIAA Computational Fluid Dynamics Conference, 2013, p. 2939.

[218] S. G. Rodrigo, F. de León-Pérez, and L. Martín-Moreno, *Extraordinary optical transmission: fundamentals and applications*, Proceedings of the IEEE, 104 (2016), pp. 2288–2306.

[219] I. Romero, J. Aizpurua, G. W. Bryant, and F. J. G. De Abajo, *Plasmons in nearly touching metallic nanoparticles: singular response in the limit of touching dimers*, Optics express, 14 (2006), pp. 9988–9999.

[220] D. V. Rovas, *Reduced-basis output bound methods for parametrized partial differential equations*, PhD thesis, Massachusetts Institute of Technology, 2003.

[221] G. Rozza, D. B. P. Huynh, and A. T. Patera, *Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations*, Archives of Computational Methods in Engineering, 15 (2008), pp. 229–275.

[222] R. Ruppin, *Extinction properties of thin metallic nanowires*, Optics communications, 190 (2001), pp. 205–209.

[223] T. Rylander, P. Ingelström, and A. Bondeson, *Computational electromagnetics*, Springer Science & Business Media, 2012.

[224] J. Saà-Seoane, *Simulation and design optimization of wave propagation in heterogeneous materials*, PhD thesis, Massachusetts Institute of Technology, 2014.

[225] J. Saà-Seoane, N. C. Nguyen, H. Men, R. Freund, and J. Peraire, *Binary optimization techniques for linear PDE-governed material design*, Applied Physics A, 109 (2012), pp. 1023–1030.

[226] A. A. Saleh and J. A. Dionne, *Toward efficient optical trapping of sub-10-nm particles with coaxial plasmonic apertures*, Nano letters, 12 (2012), pp. 5581–5586.

[227] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola, *Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index*, Computer Physics Communications, 181 (2010), pp. 259–270.

[228] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola, *Global sensitivity analysis: the primer*, John Wiley & Sons, 2008.

[229] K. J. Savage, M. M. Hawkeye, R. Esteban, A. G. Borisov, J. Aizpurua, and J. J. Baumberg, *Revealing the quantum regime in tunnelling plasmonics*, Nature, 491 (2012), pp. 574–577.

[230] D. Schmidthausler and M. Clemens, *Low-order electroquasistatic field simulations based on proper orthogonal decomposition*, IEEE Transactions on Magnetics, 48 (2012), pp. 567–570.

[231] N. Schmitt, C. Scheid, S. Lanteri, A. Moreau, and J. Viquerat, *A DGTD method for the numerical modeling of the interaction of light with nanometer scale metallic structures taking into account non-local dispersion effects*, Journal of Computational Physics, 316 (2016), pp. 396–415.

[232] J. Schöberl and S. Zaglmayr, *High order Nédélec elements with local complete sequence properties*, COMPEL-The International Journal for Computation and Mathematics in Electrical and Electronic Engineering, 24 (2005), pp. 374–384.

[233] J. A. Scholl, A. L. Koh, and J. A. Dionne, *Quantum plasmon resonances of individual metallic nanoparticles*, Nature, 483 (2012), pp. 421–427.

[234] J. A. Schuller, E. S. Barnard, W. Cai, Y. C. Jun, J. S. White, and M. L. Brongersma, *Plasmonics for extreme light concentration and manipulation*, Nature materials, 9 (2010), pp. 193–204.

[235] M. Seo, H. Park, S. Koo, D. Park, J. Kang, O. Suwal, S. Choi, P. Planken, G. Park, N. Park, et al., *Terahertz field enhancement by a metallic nano slit operating beyond the skin-depth limit*, Nature Photonics, 3 (2009), pp. 152–156.

[236] R. P. Shaw, *Boundary integral equation methods applied to wave problems.*, tech. rep., DTIC Document, 1979.

[237] L. Sirovich, *Turbulence and the dynamics of coherent structures. I-Coherent structures. II-Symmetries and transformations. III-Dynamics and scaling*, Quarterly of applied mathematics, 45 (1987), pp. 561–571.

[238] J. S. Smalley, F. Vallini, Q. Gu, and Y. Fainman, *Amplification and lasing of plasmonic modes*, Proceedings of the IEEE, 104 (2016), pp. 2323–2337.

[239] S. Smolyak, *Quadrature and interpolation formulas for tensor product of certain classes of functions*, Soviet Math. Dokl., 4 (1963), pp. 240–243.

[240] I. M. Sobol, *On the distribution of points in a cube and the approximate evaluation of integrals*, Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki, 7 (1967), pp. 784–802.

[241] I. M. Sobol, *Sensitivity estimates for nonlinear mathematical models*, Mathematical Modelling and Computational Experiments, 1 (1993), pp. 407–414.

[242] A. Sommerfeld, *Partial differential equations in physics*, vol. 1, Academic press, 1949.

[243] B. Špačková, P. Wrobel, M. Bocková, and J. Homola, *Optical biosensors based on plasmonic nanostructures: A review*, Proceedings of the IEEE, 104 (2016), pp. 2380–2408.

[244] M. Stein, *Large sample properties of simulations using Latin Hypercube Sampling*, Technometrics, 29 (1987), pp. 143–151.

[245] B. C. Stipe, T. C. Strand, C. C. Poon, H. Balamane, T. D. Boone, J. A. Katine, J.-L. Li, V. Rawat, H. Nemoto, A. Hirotsune, et al., *Magnetic recording at 1.5 Pb m^{-2} using an integrated plasmonic antenna*, Nature Photonics, 4 (2010), pp. 484–488.

[246] D. M. Sullivan, *Electromagnetic simulation using the FDTD method*, John Wiley & Sons, 2013.

[247] A. Taflove and S. C. Hagness, *Computational electrodynamics*, Artech house, 2005.

[248] M. S. Tame, K. McEnery, Ş. Özdemir, J. Lee, S. Maier, and M. Kim, *Quantum plasmonics*, Nature Physics, 9 (2013), pp. 329–340.

[249] T. V. Teperik, P. Nordlander, J. Aizpurua, and A. G. Borisov, *Quantum effects and nonlocality in strongly coupled plasmonic nanowire dimers*, Optics express, 21 (2013), pp. 27306–27325.

[250] G. Toscano, S. Raza, A.-P. Jauho, N. A. Mortensen, and M. Wubs, *Modified field enhancement and extinction by plasmonic nanowire dimers due to nonlocal response*, Optics express, 20 (2012), pp. 4176–4188.

[251] G. Toscano, J. Straubel, A. Kwiatkowski, C. Rockstuhl, F. Evers, H. Xu, N. A. Mortensen, and M. Wubs, *Resonance shifts and spill-out effects in self-consistent hydrodynamic nanoplasmonics*, Nature communications, 6 (2015).

[252] K. L. Tsakmakidis, A. D. Boardman, and O. Hess, *'Trapped rainbow' storage of light in metamaterials*, Nature, 450 (2007), pp. 397–401.

[253] P. Tsuji, B. Engquist, and L. Ying, *A sweeping preconditioner for time-harmonic Maxwell's equations with finite elements*, Journal of Computational Physics, 231 (2012), pp. 3770–3783.

[254] E. Turkel and A. Yefet, *Absorbing PML boundary layers for wave-like equations*, Applied Numerical Mathematics, 27 (1998), pp. 533–557.

[255] M. P. Ueckermann and P. F. Lermusiaux, *High-order schemes for 2d unsteady biogeochemical ocean models*, Ocean Dynamics, 60 (2010), pp. 1415–1445.

[256] K. Veroy, *Reduced-basis methods applied to problems in elasticity: Analysis and applications*, PhD thesis, Massachusetts Institute of Technology, 2003.

[257] K. Veroy and A. T. Patera, *Certified real-time solution of the parametrized steady incompressible Navier–Stokes equations: rigorous reduced-basis a posteriori error bounds*, International Journal for Numerical Methods in Fluids, 47 (2005), pp. 773–788.

[258] K. Veroy, C. Prud'homme, and A. T. Patera, *Reduced-basis approximation of the viscous Burgers equation: rigorous a posteriori error bounds*, Comptes Rendus Mathematique, 337 (2003), pp. 619–624.

[259] K. Veroy, C. Prud'homme, D. V. Rovas, and A. T. Patera, *A posteriori error bounds for reduced-basis approximation of parametrized noncoercive and nonlinear elliptic partial differential equations*, in Proceedings of the 16th AIAA Computational Fluid Dynamics Conference, vol. 3847, 2003.

[260] K. Veroy, D. V. Rovas, and A. T. Patera, *A posteriori error estimation for reduced-basis approximation of parametrized elliptic coercive partial differential equations:"convex inverse" bound conditioners*, ESAIM: Control, Optimisation and Calculus of Variations, 8 (2002), pp. 1007–1028.

[261] F. Vidal-Codina, N. C. Nguyen, M. B. Giles, and J. Peraire, *A model and variance reduction method for computing statistical outputs of stochastic elliptic partial differential equations*, Journal of Computational Physics, 297 (2015), pp. 700–720.

[262] F. Vidal-Codina, N. C. Nguyen, M. B. Giles, and J. Peraire, *An empirical interpolation and model-variance reduction method for computing statistical outputs of parametrized stochastic partial differential equations*, SIAM/ASA Journal on Uncertainty Quantification, 4 (2016), pp. 244–265.

[263] F. Vidal-Codina, J. Saà-Seoane, N. C. Nguyen, and J. Peraire, *A multiscale continuous Galerkin method for stochastic simulation and robust design of wave propagation through heterogeneous materials*, arXiv preprint arXiv:1610.00625, (2016).

[264] T. Vo-Dinh, Y. Liu, A. M. Fales, H. Ngo, H.-N. Wang, J. K. Register, H. Yuan, S. J. Norton, and G. D. Griffin, *SERS nanosensors and nanoreporters: golden opportunities in biomedical applications*, Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology, 7 (2015), pp. 17–33.

[265] 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).

[266] Z.-Y. Wang, R.-J. Zhang, H.-L. Lu, X. Chen, Y. Sun, Y. Zhang, Y.-F. Wei, J.-P. Xu, S.-Y. Wang, Y.-X. Zheng, et al., *The impact of thickness and thermal annealing on refractive index for aluminum oxide thin films deposited by atomic layer deposition*, Nanoscale research letters, 10 (2015), p. 46.

[267] D. R. Ward, N. K. Grady, C. S. Levin, N. J. Halas, Y. Wu, P. Nordlander, and D. Natelson, *Electromigrated nanoscale gaps for surface-enhanced Raman spectroscopy*, Nano letters, 7 (2007), pp. 1396–1400.

[268] J. P. Webb, *Hierarchal vector basis functions of arbitrary order for triangular and tetrahedral finite elements*, IEEE Transactions on Antennas and Propagation, 47 (1999), pp. 1244–1253.

[269] H. Whitney, *Geometric integration theory*, Courier Corporation, 2012.

[270] K. Willcox and J. Peraire, *Balanced model reduction via the proper orthogonal decomposition*, AIAA journal, 40 (2002), pp. 2323–2330.

[271] K. Willcox, J. Peraire, and J. D. Paduano, *Application of model order reduction to compressor aeroelastic models*, Journal of Engineering for Gas Turbines and Power (Transactions of the ASME), 124 (2002), pp. 332–339.

[272] C. Williams, M. Misra, S. Andrews, S. Maier, S. Carretero-Palacios, S. Rodrigo, F. Garcia-Vidal, and L. Martin-Moreno, *Dual band terahertz waveguiding on a planar metal surface patterned with annular holes*, Applied Physics Letters, 96 (2010), p. 011101.

[273] H. Xu, E. J. Bjerneld, J. Aizpurua, P. Apell, L. Gunnarsson, S. Petronis, B. Kasemo, C. Larsson, F. Hook, and M. Kall, *Interparticle coupling effects in surface-enhanced Raman scattering*, in BiOS 2001 The International Symposium on Biomedical Optics, International Society for Optics and Photonics, 2001, pp. 35–42.

[274] H. Xu, E. J. Bjerneld, M. Käll, and L. Börjesson, *Spectroscopy of single hemoglobin molecules by surface enhanced Raman scattering*, Physical review letters, 83 (1999), p. 4357.

[275] K. Yasumoto, *Electromagnetic theory and applications for photonic crystals*, CRC press, 2005.

[276] K. S. Yee, *Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media*, IEEE Trans. Antennas Propag, 14 (1966), pp. 302–307.

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

[278] Y. Zhang, Y.-W. Tan, H. L. Stormer, and P. Kim, *Experimental observation of the quantum Hall effect and Berry's phase in graphene*, Nature, 438 (2005), pp. 201–204.

[279] W. Zhu, R. Esteban, A. G. Borisov, J. J. Baumberg, P. Nordlander, H. J. Lezec, J. Aizpurua, and K. B. Crozier, *Quantum mechanical effects in plasmonic structures with subnanometre gaps*, Nature communications, 7 (2016).

[280] P. Zijlstra, J. W. Chon, and M. Gu, *Five-dimensional optical recording mediated by surface plasmons in gold nanorods*, Nature, 459 (2009), p. 410.

[281] J. Zuloaga, E. Prodan, and P. Nordlander, *Quantum description of the plasmon resonances of a nanoparticle dimer*, Nano letters, 9 (2009), pp. 887–891.