Inverse Design of Random Emitters in Nanophotonics

by

Wenjie Yao

B.S., Peking University (2016)
S.M., Massachusetts Institute of Technology (2020)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2022

Author . . . .
Department of Electrical Engineering and Computer Science
August 10, 2022

Certified by. . . . .
Steven G. Johnson
Professor of Applied Mathematics and Physics
Thesis Supervisor

Accepted by . . . .
Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

Inverse Design of Random Emitters in Nanophotonics

by

Wenjie Yao

Submitted to the Department of Electrical Engineering and Computer Science
on August 10, 2022, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Electrical Engineering and Computer Science

Abstract

Incoherent light from random emitters, such as thermal radiation, are very common in nature. However, modeling such random emitters may be challenging, as it naively requires Maxwell's equations to be solved for all emitters to obtain the total response, which becomes computationally intractable in conjunction with large-scale optimization (inverse design). In this work, we present a trace formulation of random emitters that can be efficiently combined with inverse design, even for topology optimization over thousands of design degrees of freedom.

We begin with a trivial case where the emitter is at a single location with random orientations, which leads to compute the local density of states (LDOS). In a previous work, a shape-independent upper limit was derived for LDOS, but simple geometries such as bowtie are 2-3 orders of magnitude away from this limit. By computational optimization of air-void cavities in metallic substrates, we show that the LDOS can reach within a factor of ≈ 10 of the upper limits, and within a factor ≈ 4 for the single-polarization LDOS, demonstrating that the theoretical limits are nearly attainable.

We then study the more general case where emitters are distributed randomly in space. We present several examples of incoherent-emission topology optimization (TopOpt), including tailoring the geometry of fluorescent particles, a periodically emitting surface, and a structure emitting into a waveguide mode.

Finally, we employ our trace formulation for inverse design of nanopatterned surfaces that maximize spatially averaged surface-enhanced Raman (SERS) spectra from molecules distributed randomly throughout a material or fluid. This leads to radically different designs than optimizing SERS emission at a single known location, as we illustrate using several 2D design problems addressing effects of hot-spot density, angular selectivity, and nonlinear damage. We obtain optimized structures that perform about $4\times$ better than coating with optimized spheres or bowtie structures and about $20\times$ better when the nonlinear damage effects are included.

Thesis Supervisor: Steven G. Johnson
Title: Professor of Applied Mathematics and Physics

Acknowledgments

I would like to acknowledge and give my warmest thanks to my supervisor (Steven G. Johnson) who made this work possible. His guidance and advice carried me through all the stages of writing my project. I would also like to give thanks to Mohammed Benzaouia, Owen Miller, Francesc Verdugo, Rasmus E. Christiansen, and Henry O. Everitt, who provided me with all kinds of help during my program. Finally, I would like to give thanks to my family members (parents and brothers) who have been supporting me all the time.

Contents

List of Figures

11

Chapter 1

Introduction

Incoherent light from random emitters is everywhere. For example, when one heats up material, thermal vibrations of the particles can be treated as random emitters that produce thermal radiation in both far [18] and near [10,127] fields. Fluorescence and spontaneous emission are produced by excited atoms randomly dropping down from the excited states and emitting a photon, which can also be modeled as random emitters [64,96,119]. There are many other phenomena that involve random emitters such as scintillation [16,130], Casimir and van der Waals forces [38], Raman scattering in fluid suspensions [117,118], incoherent plane waves [114,153] (which can be transformed to random sources via the equivalence principle [45]), and even scattering from surface roughness via a Born approximation [57]. As a matter of fact, coherent light sources such as lasers are primarily constructed artificially [31], while most light sources in nature are incoherent.

However, accurate modeling of such spatially random emitters can pose severe computational challenges, because a direct approach would involve averaging the results of many simulations over an ensemble of sources [8,84,126].For example, to evaluate the total power from randomly distributed spontaneous emission particles, the direct approach would be computing the power from one particle at a time and then sum them up, which requires a number of simulations that equals to the number of particles (typically a huge number). This challenge is compounded when one wishes to perform *inverse design* [98]—large-scale optimization of emission over many

geometric parameters, perhaps even over "every pixel" of a design region via topology optimization (TopOpt) [53]—because one must then repeat the computation 10s–1000s of times as the design evolves, e.g. to maximize spontaneous emission [81, 129, 150, 159] or Raman emission [22] from a single molecule, much less a distribution of sources. Therefore, it is critical to find a method that can efficiently compute the total (average) response from random emitters as well as optimize it.

1.1 Overview

In Chap. 2, we present a unified framework for computing incoherent emission that can be efficiently combined with inverse design, even for TopOpt over thousands of design degrees of freedom. We show that for the general case with many input/output channels, the emission can be computed by a small number of "eigen-sources" of a Hermitian operator, which can be determined by a Rayleigh-quotient optimization [80] that is *combined* with the inverse-design (geometric) optimization. In the special case of emission into a small number K of channels, such as K far-field directions, K waveguide modes, or K points in space, we show that a simple algebraic manipulation transforms the problem into K simulations —this unifies and generalizes known results based on Kirchhoff's law of thermal radiation [39, 124] or (more generally) reciprocity [52, 130] for computing emission into a single planewave direction—but our alternative approach yields a small number of solves even for $K \to \infty$. The other well known special case is that of a single emitter location with random orientations, which reduces to the local density of states (LDOS) via three Maxwell solves [96, 108], and this appears as another low-rank special case in our formulation.

In Chap. 3, we study the trivial case where the emitter is at a *single* location in space but with random polarization, which leads to the electric LDOS [61]. One interesting question is how close can we approach the upper limit of LDOS proposed in a recent work [95], as conventional simple resonant structures (such as bowtie and nanorod antennas) were often orders of magnitude from this upper limit. We perform many-parameter shape optimization of LDOS for cavities formed by voids

in silver (since it has the largest bounds) at wavelengths λ from 400–900 nm and a $d = 50$ nm emitter–metal separation. We obtain single-polarization LDOS values within a factor of ≈ 4 of the theoretical upper bounds, and total (all-polarization) LDOS within a factor of ≈ 10 of the bounds. Of course, real cavities would have a finite thickness of metal, but our goal is to attain the maximum possible LDOS—we find that a finite-thickness coating has slightly worse performance, but $> 95\%$ of the LDOS of the infinite metal is attained by ≈ 100 nm thickness at $\lambda = 500$ nm, and more generally we can theoretically bound [95] the improvement attainable with *any* additional structure of air voids outside of our cavity. Although our focus is on fundamental upper limits rather than manufacturable cavities, we find that simple shapes (optimized cylinders) are within $\approx 20\%$ of the LDOS of optimized many-parameter irregular shapes, analogous to results we obtained previously for optimized scattering and absorption [95]. (Our optimized cavities are deeply subwavelength along their shortest axes, very different from the non-plasmonic resonant regime where the diameter is much larger than the skin depth so that the walls simply act as mirrors.) Moreover, we find that optimizing for a single emitter polarization (the "polarized" LDOS) does nearly as well (within $\approx 10\%$) as optimizing the total LDOS (power summed over all emitter polarizations), reminiscent of earlier results in 2D dielectric cavities where LDOS optimization arbitrarily picked one polarization to enhance [81]. Although it is possible that even tighter LDOS bounds could be obtained in future results by incorporating additional physical constraints [40, 71, 97, 99], we believe that our results show that the existing bounds are already closely related to attainable performance and provide useful guidance for optical cavity design.

In Chap. 4, we briefly review the key ideas in density-based TopOpt [53] and show how the trace formulation can be applied to perform density-based TopOpt on several example problems in 2D: fluorescence from an optimized nanoparticle, enhanced emission from a corrugated surface analogous to a light-emitting diode [32, 52], and optimized emission into a waveguide. In each case, the emission is not from a single molecule, but the average power produced by an ensemble of incoherent emitters at every point in some material. Those problems are computationally expensive to

model in convention. However, with the trace formulation, we show that only a few Maxwell's solves are required to compute the average objective in each optimization step, which greatly reduces the computation costs.

In Chap. 5, we propose an efficient technique for simulation and inverse design of spatially averaged Raman enhancement, building from the trace formulation, and analyze its results for TopOpt applied to several example problems addressing effects of hot-spot density, angular selectivity, and nonlinear damage. In particular, we show that spatially averaged Raman emission in a single direction can be modeled with only two Maxwell simulations, one for the pump process and the other a "reciprocal" solve for the average emission over *all* molecule locations, easily generalized to support nonlinear damage and/or anisotropic Raman polarizability. Moreover, this formulation is straightforwardly compatible with large-scale inverse design, requiring only two additional "adjoint" simulations [98] to compute the sensitivity of the output power with respect to "every pixel" of the design (e.g. a material density at every point in TopOpt). Previous authors employed a simplified $\int |\mathbf{E}_1|^4$ metric for distributed Raman emission [120, 137], where $\mathbf{E}_1$ is the pump electric field, and we show that this is a special case of our framework when the emission is in the same direction as the pump, the Raman shift is negligible ($\omega_1 \approx \omega_2$), and the Raman molecule is isotropic. We also analyze how the $\sim |\mathbf{E}_1|^4$ nonlinearity favors hot spots and field singularities (from sharp corners) in 3D, but less so in 2D (where the singularities have a finite integral). We apply TopOpt to various example problems in 2D to illustrate the key tradeoffs and physical effects: normal incidence and emission, 30° pump and normal emission (which performs nearly as well but with a very different design), emission with UV-like nonlinear damage (again leading to very different designs), and Raman emission only from a material-surface coating [89] rather than in a volume/fluid coating. By comparing TopOpt for periodic surfaces of varying periods, we observe an optimal density of the resulting hot spots. We obtain optimized structures that perform about 4× better than coating with optimized spheres or bowtie structures and about 20× better when the nonlinear damage effects are considered. Also, we find similar optimal structures when Raman molecules are distributed only on the metal

surface, as opposed to throughout a volumetric (fluid) coating.

1.2   Numerical tools

Throughout this work, we use two main numerical tools for simulating the electro-magnetic processes: SCUFF [122] and Gridap [6].

In Chap. 3 where we optimize the LDOS inside an metallic cavity, we use SCUFF for the boundary element method (BEM) implementation. This is more efficient than volume-based discretizations since the air-void cavity can simply be described by the separation surface between the air and metal and all the unknowns are on the surface in a BEM scheme. SCUFF is a free, open-source software implementation of the BEM of electromagnetic scattering, more details about this implementation (including some validation examples) can be found in my master thesis [158].

In Chap. 4 and Chap. 5 where we implement TopOpt for inverse design, we employ the Gridap.jl package [6] (in the Julia language [13]) for the finite element method (FEM). Unlike BEM, one must discretize not only the boundaries but also the whole computation domain, and we use Gmsh [36] to generate this mesh. Gridap.jl is also a free, open-source FEM package, which allows us to efficiently code highly customized FEM-based trace formulations in a high-level language, with the construction of the adjoint problem aided by automatic-differentiation (AD) tools [49, 125]. Because we collaborated directly with Gridap author Francesc Verdugo to develop Gridap's capa-bilities for electromagnetic design, we contributed the first tutorials on this topic to the Gridap documentation. In `https://gridap.github.io/Tutorials/dev/pages/t012_emscatter/`, we provide a tutorial on a electromagnetic scattering problem in 2D, explaining the basic usage of Gridap for electromagnetic scattering prob-lems. Furthermore, in `https://gridap.github.io/Tutorials/dev/pages/t018_TopOptEMFocus/`, we provide a second tutorial on a simple inverse design problem of electromagnetics, showing how density-based topology optimization can be imple-mented with Gridap.

Chapter 2

Trace formulation for photonic inverse design with random emitters

In this chapter, we present a unified framework for inverse design of incoherent emission, combining a *trace formulation* adapted from recent work [119, 123, 127] with a new algorithm to *simultaneously* optimize the geometry and evolve to an accurate estimate of the average emission/trace.

To begin with, we review the formulation of the frequency-domain Maxwell equations as a linear equation, discretized for numerical computation, with physical quantities like power as quadratic forms (Sec. 2.1). Then we show how the ensemble average of such an expression over a distribution of random current sources can be rewritten as a deterministic trace formula (Sec. 2.2). Next, we derive the expression for the correlation matrix which is a key quantity for trace formulation (Sec. 2.3). The objective trace formula is presented in both factorization form (Sec. 2.4) and factorization-free form (Sec. 2.8). Finally, we explain how such a trace formula can be evaluated efficiently in the context of photonics optimization, both in the "easy" cases of coupling to a small number of output/input channels (Sec. 2.5 and Sec. 2.6) as well as in the more general cases of a continuum of outputs (Sec. 2.7).

2.1 Wave sources and quadratic outputs

In the frequency domain, the linear Maxwell equations for the electric field $\mathbf{E}$ in response to a time-harmonic current source at a frequency ω are [54]

$$\left[\nabla \times \frac{1}{\mu} \nabla \times - \left(\frac{\omega}{c} \right)^2 \varepsilon \right] \mathbf{E} = \mathbf{f} \,, \tag{2.1}$$

where $\varepsilon(\mathbf{x}, \omega)$ is the relative electric permittivity, μ is the relative magnetic permeability ($\mu \approx 1$ for most materials at optical and infrared wavelengths, so we assume $\mu = 1$ throughout this work), c is the speed of light in vacuum, and $\mathbf{f} = i\omega \mathbf{J}$ is a current-source term.

Numerically, one discretizes the problem (e.g. using finite elements [54]) into a linear equation:

$$A\mathbf{u} = \mathbf{b} \,, \tag{2.2}$$

where A is a matrix representing the Maxwell operator on the left-hand of Eq. (2.1), $\mathbf{u}$ is a vector representing the discretized electric (and/or magnetic) field, and $\mathbf{b}$ is a vector representing the discretized source term. In the following, it is algebraically convenient to work with such a discretized (finite-dimensional) form, to avoid cumbersome infinite-dimensional linear algebra, but one could straightforwardly translate to the latter context as well [55].

Most physical quantities P of interest in photonics—such as power (via the Poynting flux), energy density, and force (via the Maxwell stress tensor)—can be expressed as quadratic functions of the electromagnetic fields $\mathbf{u}$. Since these are real-valued quantities, they correspond in particular to Hermitian quadratic forms

$$P = \mathbf{u}^\dagger O \mathbf{u} \,, \tag{2.3}$$

where $\dagger$ denotes the conjugate transpose (adjoint) and $O = O^\dagger$ is a Hermitian matrix/operator. In this work, we are mainly concerned with computing emitted *power* P, which is constrained by the outgoing boundary conditions to be non-negative, in which case O must furthermore be a *positive semi-definite* Hermitian matrix (i.e., non-

negative eigenvalues) in the subspace of permissible $\mathbf{u}$, a property that will be useful in Sec. 2.7.

2.2 Trace formula for random sources

Now, consider the case where one has an ensemble of random current sources $\mathbf{b}$ drawn from some statistical distribution with zero mean and a known correlation function (e.g. a known mean-square current at each point if they are spatially uncorrelated). In this case, we wish to compute the ensemble average, denoted by $\langle \cdots \rangle$, of our quadratic form Eq. (2.3):

$$\langle P \rangle = \langle \mathbf{u}^\dagger O \mathbf{u} \rangle = \langle \mathbf{b}^\dagger A^{-\dagger} O A^{-1} \mathbf{b} \rangle \,, \tag{2.4}$$

where $A^{-\dagger}$ denotes $(A^{-1})^\dagger = (A^\dagger)^{-1}$. Note that only $\mathbf{b}$ is random in the right-hand expression.

Naively, this average could be computed by a brute-force method in which one explicitly solves the Maxwell equations ($\mathbf{u} = \mathbf{A}^{-1}\mathbf{b}$) for many possible sources $\mathbf{b}$ and then integrates over the distribution, perhaps by a Monte-Carlo (random-sampling) method. That approach is possible, and has been accomplished e.g. for evaluating thermal radiation [84, 126], but is computationally expensive. Worse, such a direct approach quickly becomes prohibitive in the context of inverse design, where the averaging must be repeated for many geometries over the course of solving an optimization problem using an iterative algorithm.

Instead, we adapt "trace formula" techniques that have been developed for similar problems in thermal radiation [127] and spontaneous emission [119], where one must compute the average effect of many random current sources distributed throughout a volume. The basic trick (as reviewed in yet another related setting in [123]) is to write the scalar $\langle P \rangle$ as a 1×1 "matrix" trace, and then employ the cyclic-shift property [76] to group the $\mathbf{b}$ terms together:

$$\langle P \rangle = \langle \mathbf{b}^\dagger A^{-\dagger} O A^{-1} \mathbf{b} \rangle = \operatorname{tr} \langle \mathbf{b}^\dagger A^{-\dagger} O A^{-1} \mathbf{b} \rangle = \operatorname{tr} \left[A^{-\dagger} O A^{-1} \langle \mathbf{b}\mathbf{b}^\dagger \rangle \right] \,. \tag{2.5}$$

Here, the ensemble average is now confined to the $\langle \mathbf{b}\mathbf{b}^\dagger \rangle$ term, which is just the correlation matrix B [56] of the currents; such a matrix is positive semi-definite, so it can be factorized [143] (for convenience below) as

$$\langle \mathbf{b}\mathbf{b}^\dagger \rangle = B = DD^\dagger, \tag{2.6}$$

for some known matrix D. (For the case of finite-element discretizations, we show that B is a sparse matrix that is straightforward to assemble and D is, for example, a sparse Cholesky factor [27].)

2.3 Correlation matrix

One can express the frequency-domain Maxwell equations either in terms of the electric field $\mathbf{E}$, in which case the source term is proportional to $\mathbf{J}$, or in terms of the magnetic field $\mathbf{H}$, in which case the source term is proportional to $\nabla \times \mathbf{J}$ [54]. These two formulations lead to different B correlation matrices.

In particular, we consider the case where the currents $\mathbf{J}$ (at a frequency ω) are spatially uncorrelated with a given correlation function:

$$\langle \mathbf{J}(\mathbf{x})\mathbf{J}(\mathbf{x}')^\dagger \rangle = C(\mathbf{x})\delta(\mathbf{x} - \mathbf{x}'), \tag{2.7}$$

where C is a given 3×3 Hermitian positive-semidefinite correlation matrix. For example, in 2D with in-plane electric currents, one has

$$C = \begin{pmatrix} J_0^2 & & \\ & J_0^2 & \\ & & 0 \end{pmatrix}, \tag{2.8}$$

where $J_0^2(\mathbf{x})$ is the mean-square current at $\mathbf{x}$. For isotropic random currents, $C = J_0^2 I$ where I is the identity matrix.

In a finite-element method, the source vector $\mathbf{b}$ is constructed by taking inner

products of the source current with real vector-valued basis "element" functions $\hat{\mathbf{v}}_n$ (Nedelec elements in 3D, or $\hat{v}_n\hat{\mathbf{z}}$ with scalar Lagrange elements $\hat{v}_n$ in 2D for z-polarized fields) [54]. That is, the components of $\mathbf{b}$ are

$$b_n = \int \hat{\mathbf{v}}_n \cdot (\text{source current})\, \mathrm{d}\Omega. \tag{2.9}$$

For an electric-field formulation with a source current $\mathbf{J}$, we obtain the correlation function:

$$
\begin{aligned}
B_{mn} &= \langle b_m b_n^* \rangle, \\
&= \left\langle \iint \hat{\mathbf{v}}_m(\mathbf{x})^T \mathbf{J}(\mathbf{x}) \mathbf{J}(\mathbf{x}')^\dagger \hat{\mathbf{v}}_n(\mathbf{x})\, \mathrm{d}\Omega \mathrm{d}\Omega' \right\rangle, \\
&= \iint \hat{\mathbf{v}}_m(\mathbf{x})^T \left\langle \mathbf{J}(\mathbf{x}) \mathbf{J}(\mathbf{x}\prime)^\dagger \right\rangle \hat{\mathbf{v}}_n(\mathbf{x}')\, \mathrm{d}\Omega \mathrm{d}\Omega', \\
&= \int \hat{\mathbf{v}}_m^T C \hat{\mathbf{v}}_n\, \mathrm{d}\Omega.
\end{aligned} \tag{2.10}
$$

For localized basis functions (as in a finite-element method), this results in an extremely sparse matrix B—it is zero if $\hat{\mathbf{v}}_m$ and $\hat{\mathbf{v}}_n$ don't overlap, or in regions where the mean-square current C is zero. (If C is the identity, B is equal to the Gram matrix of the basis.) Note also that, by construction, B is a Hermitian semidefinite matrix, so it has factorization $B = DD^\dagger$, such as a Cholesky factorization [143].

For a magnetic-field formulation, $\mathbf{J}$ is replaced by $\nabla \times \mathbf{J}$ above, but we can simply integrate by parts [55] to move the $\nabla\times$ curl operation to act on the basis functions, yielding:

$$B_{mn} = \langle b_m b_n^* \rangle = \int (\nabla \times \hat{\mathbf{v}}_m)^T C (\nabla \times \hat{\mathbf{v}}_n)\, \mathrm{d}\Omega. \tag{2.11}$$

Again, this yields a sparse Hermitian semidefinite matrix B.

2.4 Trace formulation with factorization

Algebraically, expressing our results in terms of D below leads to convenient Hermitian matrices, but we show in Sec. 2.8 that the final algorithms can easily employ B

directly to avoid the computational cost of an explicit factorization. In the simple case where random currents are *spatially uncorrelated*, which holds for spontaneous emission and thermal emission in local materials [74], B and D are conceptually *diagonal* linear operators whose diagonal entries are the mean-square and root-mean-square currents, respectively, at each point in space. Whether this leads to a strictly diagonal *matrix* depends on the discretization scheme as explained in Sec. 2.3. For instance, in the case of thermal and quantum fluctuations, the mean-square currents are given by the fluctuation–dissipation theorem (FDT) [74], while for spontaneous emission one can use the FDT with a "negative temperature" determined by the population inversion [112, 116]. The FDT tells us that the random currents are spatially uncorrelated as long as the materials can be described by local susceptibilities, e.g. a local ε and μ—although this macroscopic description in principle breaks down at atomic lengthscales, or even at few-nm lengthscales in metals experiencing nonlocal ballistic charge transport [121], in practice electromagnetic modeling with local materials has proved extremely accurate for describing almost all experimental circumstances. (Note that the lack of spatial correlations is a property of the *source currents*, which do not interact in the linear Maxwell equations, not of the resulting material polarizations and electromagnetic fields, which *are* spatially correlated [128, Sec. 4.1.2].)

Inserting Eq. (2.6) into Eq. (2.5), we obtain our objective as the trace of a deterministic Hermitian matrix H (which is positive-semidefinite if O is, as for power), given by:

$$\langle P \rangle = \text{tr} \underbrace{\left[(A^{-1}D)^\dagger O (A^{-1}D) \right]}_{H} . \tag{2.12}$$

The challenge now is to efficiently compute such a matrix trace. Evaluating a trace is easy once the matrix elements are known—it is the sum of the diagonal entries— but the difficulty in Eq. (2.12) is the computation of $A^{-1}D$. Recall that the $N \times N$ matrix A is a discretized Maxwell operator where N is the number of grid points (or basis functions), a huge matrix (especially in 3D). There are fast methods to solve for $A^{-1}(D\mathbf{v})$ for any *single* right-hand side $\mathbf{v}$, typically because the matrix A is *sparse* (mostly zero) as in finite-element methods [54], but computing the *whole*

matrix $A^{-1}D$ corresponds to solving N right-hand sides. Equivalently, computing explicit (dense) matrix inverses A^{-1} is typically prohibitively expensive (in both time and storage) for matrices arising in large physical systems [27]. Fortunately, a large number of "iterative" algorithms have been proposed for *estimating* matrix traces to any desired accuracy using relatively *few* matrix–vector products [47, 145], and what remains is to find a method well-suited to inverse design.

2.5   Trace computation: Few output channels

In the important special cases where the desired output is the power in a small number (K) of discrete directions/channels/ports, or perhaps the intensity at a few points in space, we show in this section that the trace computation Eq. (2.12) simplifies to only K scattering problems. This fact is a generalization of earlier results commonly derived from electromagnetic reciprocity [20], such as the well-known Kirchhoff's law of thermal radiation (reciprocity of emission and absorption) [124] or analogous results for scintillation [130]. More generally, this simplification arises whenever the matrix O in Eq. (2.3) is *low rank*.

For example, suppose that the objective function is the electric field intensity $\|\mathbf{E}(\mathbf{x}_1)\|^2$ at a single point $\mathbf{x}_1$ in space, which is the case for "metalens" optimization problems in which one is maximizing intensity at a focal spot [11]. In matrix notation for a discretized problem, this quantity corresponds to

$$P = \|\mathbf{E}(\mathbf{x}_1)\|^2 = \|\mathbf{e}_1^\dagger \mathbf{u}\|^2 = \mathbf{u}^\dagger \underbrace{\mathbf{e}_1 \mathbf{e}_1^\dagger}_{O} \mathbf{u} \,, \tag{2.13}$$

where $\mathbf{e}_1$ is the unit vector with a nonzero entry at the location ("grid point") corresponding to $\mathbf{x}_1$. We then have a rank-1 [76] matrix $O = \mathbf{e}_1 \mathbf{e}_1^\dagger$, and the trace Eq. (2.12) simplifies to $\langle P \rangle = \mathbf{v}_1^* \mathbf{v}_1 = \|\mathbf{v}_1\|^2$ where

$$\mathbf{v}_1 = D^\dagger A^{-\dagger} \mathbf{e}_1 \tag{2.14}$$

and $A^{-\dagger}\mathbf{e}_1$ corresponds to solving a (conjugate-) *transposed* Maxwell problem with a "source" $\mathbf{e}_1$ at the *output* location, which is closely related to electromagnetic reciprocity [20].

Another important example where O is low-rank arises when the output P is the power in one (or more) orthogonal "wave channels" [136], such as waveguide modes, planewave directions (e.g. diffraction orders), or spherical waves. In such cases the power in a given channel can be computed by squaring a *mode-overlap integral* (e.g. a Fourier component for planewaves) of the form $\|\mathbf{o}_1^\dagger\mathbf{u}\|^2$ [136]. Exactly as in the single-point case above, this corresponds to a rank-1 matrix $O = \mathbf{o}_1\mathbf{o}_1^\dagger$ and one must solve only a single "reciprocal" scattering problem to obtain the trace, where the "source" term is the (conjugated) *output* mode $\mathbf{o}_1$. This is precisely the situation in Kirchhoff's law, where in order to compute the average thermal radiation (emissivity) in a given direction, one solves a reciprocal problem for the absorption of an incident planewave in the opposite direction (the absorptivity) [39, 52, 124]. A similar technique was recently applied to optimize the average power emitted in the normal direction from a scintillation device [130].

More generally, such cases correspond to an output quadratic form O that takes a low-rank [76] form:

$$O = \sum_{i=1}^{K} \mathbf{o}_i\mathbf{o}_i^\dagger, \tag{2.15}$$

where K is the number of rank-1 terms $\mathbf{o}_i\mathbf{o}_i^\dagger$ (e.g. output channels/ports, output points, or other "overlap integrals"). Substituting Eq. (2.15) into Eq. (2.12) and applying the cyclic-trace identity, we obtain:

$$
\begin{aligned}
\langle P \rangle &= \sum_{i=1}^{K} \operatorname{tr}\left[(A^{-1}D)^\dagger \mathbf{o}_i\mathbf{o}_i^\dagger A^{-1}D\right] \\
&= \sum_{i=1}^{K} \mathbf{o}_i^\dagger A^{-1}D(A^{-1}D)^\dagger \mathbf{o}_i \\
&= \sum_{i=1}^{K} \mathbf{v}_i^\dagger\mathbf{v}_i = \sum_{i=1}^{K} \|\mathbf{v}_i\|^2,
\end{aligned}
\tag{2.16}
$$

where

$$\mathbf{v}_i = D^\dagger A^{-\dagger} \mathbf{o}_i \tag{2.17}$$

corresponds to a single "reciprocal" Maxwell solve $A^{-\dagger} \mathbf{o}_i = (A^{-T} \mathbf{o}_i^*)^*$ (a single scattering problem) for each i. (Electromagnetic reciprocity simply corresponds to the fact that $A^T = A$ for reciprocal materials [20].) Hence, the full trace—the average emission into K channels—can be computed with only K solves, and in many such cases $K = 1$.

2.6   Trace computation: Few input channels

One trivial special case in which the trace computation drastically simplifies is that of only a few sources or a few input channels, most famously in the case of the local density of states (LDOS): emission by a molecule at a *single* location in space but with a random polarization [96,108]. In the case of LDOS, this reduces the trace computation to three Maxwell solves, one per principal polarization direction, making the problem directly tractable for topology optimization [81,150,159]. More generally, this situation corresponds to the correlation matrix B being low-rank: if B is rank K, we can compute the trace in K solves.

In particular, suppose that the currents $\mathbf{b}$ are of the form $\mathbf{b} = \sum_{i=1}^K \beta_i \mathbf{b}_i$ where the $\mathbf{b}_i$ are "input channel" basis functions (e.g. a point source with a particular orientation, or an equivalent-current source for a waveguide mode [108]) and β_i are uncorrelated random numbers with zero mean and unit mean-square. Then the correlation matrix $B = \langle \mathbf{b}\mathbf{b}^\dagger \rangle$ is simply the rank-K matrix $B = \sum_i \mathbf{b}_i \mathbf{b}_i^\dagger$. In this case, the trace simplifies to:

$$\operatorname{tr} H = \sum_{i=1}^K \mathbf{u}_i^\dagger O \mathbf{u}_i \ , \tag{2.18}$$

where computing $\mathbf{u}_i = A^{-1} \mathbf{b}_i$ again requires only K solves, one per source $\mathbf{b}_i$.

2.7 Trace computation: Many output channels

In general, neither the matrix O nor the matrix B are low rank—for example, one may
be interested in the total power radiated into a *continuum* of angles above a surface, or
some other infinite set of possible far-field distributions, from sources distributed over
a continuous spatial region. Fortunately, it turns out that there is another structure
we can exploit: the Hermitian matrix $H = (A^{-1}D)^\dagger O(A^{-1}D)$ from Eq. (2.12) is itself
typically *approximately low rank* ("numerically low rank" [85]) even if O is not: the
trace, which equals the sum of the eigenvalues of H [76], is *dominated by a few of H's
largest eigenvalues*. In this section, we first explain why that is the case, and then
show how it can be exploited to efficiently estimate the trace during optimization.

There are two reasons to expect approximate low-rank structure of H. First, on
physical grounds, emission enhancement arises due to resonances (via the Purcell ef-
fect) [2], but in any finite volume there is some limit to the number of resonances that
can interact strongly with emitters in a given bandwidth, related to an average den-
sity of states [160]. The traditional definition of resonant modes corresponds to poles
of A^{-1} at complex resonant frequencies, which are (linear or nonlinear) eigenvalues ω
satisfying $\det A(\omega) = 0$ [105]; analogously, Eq. (2.12) decomposes the total power into
a sum of eigenvalues corresponding to "resonant current" sources which diagonalize H
at a given frequency. More explicitly, if $A^{-1}D$ can be accurately approximated by the
action of K resonances of A (a quasinormal mode expansion [35, 72]), so that A^{-1} can
be replaced by a rank-K matrix, it follows that H is also approximately rank $\leq K$
(since it is a product of rank-deficient matrices [76]). Moreover, geometric optimiza-
tion to maximize the emitted power modifies the structure to further enhance one or
more resonances [81], and we observe that this sometimes increases the concentration
of the trace into a few eigenvalues of H; that is, optimized structures tend to be even
lower rank. Second, in a more general mathematical sense, the matrix H is built
from *off-diagonal blocks* of the Green's function matrix A^{-1}, connecting sources (at
the the support of D) to emitted power at some other location (the support of O,
e.g. where the Poynting flux is computed), and off-diagonal blocks of Green's functions

are known to be approximately low-rank [41]. This is closely related to fast methods for integral equations, such as the fast-multipole method and others [37]; essentially, far fields mostly depend on low-order spatial moments of the near fields/currents.

If $\operatorname{tr} H$ is dominated by $K \ll N$ largest eigenvalues of the $N \times N$ matrix H, then one merely needs a numerical algorithm to compute the K *extremal* (largest-magnitude) eigenvalues using only a sequence matrix–vector products $H\mathbf{v}$ (corresponding to individual scattering problems). Fortunately, there are many such algorithms, especially for Hermitian H [68, 73], and one can simply increase K until the trace converges to any desired tolerance. We argue here that methods based on Rayleigh-quotient maximization are particularly attractive for inverse design because they can be *combined* with geometric/topology optimization. The key fact is that one can express the sum of the largest K eigenvalues as the maximum of a block Rayleigh quotient [58, 68, 69, 80], and for positive semidefinite H (= positive semidefinite O) this sum is a lower bound on the trace [69]:

$$\operatorname{tr} H \geq \max_{V \in \mathbb{C}^{N \times K}} \operatorname{tr} \left[(A^{-1}DV)^{\dagger} O (A^{-1}DV)(V^{\dagger}V)^{-1} \right] , \qquad (2.19)$$

where V represents any K-dimensional subspace basis, so that one is maximizing the trace over all possible subspaces. This $\geq$ becomes equality for $N = K$, but in many problems (below) we find that $K < 10$ suffices for $< 1\%$ error in the trace (and, as expected from the arguments above, we find that the required K increases with the diameter of the emission region).

Computationally, one can maximize the right-hand side of Eq. (2.19) by some form of gradient ascent [68, 80], each step of which only requires the evaluation of $A^{-1}DV$ for a $N \times K$ matrix V. That is to say, one only needs K Maxwell solves at each step (instead of N for the full matrix H), which vastly reduces the computational cost.

Moreover, this Rayleigh-quotient maximization formula is especially attractive in the context of inverse design, because it can be *combined with the geometric optimization* itself. That is, instead of "nesting" the trace computation inside a larger geometric optimization procedure, we can simply add V to the geometry degrees of

freedom and optimize over both V and the geometry *simultaneously.* The full inverse-design problem with incoherent emission can now be bounded by a *single* optimization problem:

$$\langle P \rangle_{\text{optimum}} \geq \max_{\text{geometry}, V \in \mathbb{C}^{N \times K}} \text{tr} \left[(A^{-1}DV)^{\dagger} O (A^{-1}DV)(V^{\dagger}V)^{-1} \right] , \qquad (2.20)$$

where the geometric parameters (e.g. material densities [53] or level sets [147]) only affect A and (perhaps) D, and may be subject to some geometric and/or material constraints. The gradient of the right-hand side with respect to the geometry can be computed efficiently with adjoint methods [98], whereas the gradient with respect to V has a simple analytical formula [58], so a variety of gradient-based optimization algorithms [21] can be applied to simultaneously evolve both V and the geometry. Furthermore, the Rayleigh quotient has the nice property that, since we are maximizing a lower bound on the full trace, the actual performance $\langle P \rangle$ is guaranteed to be at least as good as the estimated performance at every optimization step.

2.8 Factorization-free trace formulation

Although it is conceptually attractive to use a trace formulation Eq. (2.12) in terms of the Hermitian matrix H, this formulation required a factorization $B = DD^{\dagger}$ of the correlation matrix B. Computationally, it is desirable to avoid this factorization, especially if the current distribution (and hence B) depends on the geometric degrees of freedom ρ (which would require us to differentiate through the matrix factorization in our adjoint calculation). Instead, it is straightforward to reformulate our optimization problems Eq. (2.16) and Eq. (2.20) in terms of B alone using a change of variables.

For the few-output-channel case in Sec. 2.5, one can simply start with Eq. (2.12) and rewrite it as $\langle P \rangle = \text{tr}[A^{-\dagger}OA^{-1}B]$, which for a low-rank O simplifies, similar to Eq. (2.16), to

$$\langle P \rangle = \sum_{i=1}^{K} \mathbf{u}_i^{\dagger} B \mathbf{u}_i , \qquad (2.21)$$

where $A^{\dagger}\mathbf{u}_i = \mathbf{o}_i$.

For the many-channel case of Eq. (2.20), the key point is that we can choose V to be orthogonal to the nullspace $N(D)$ of D, as any nullspace component would contribute nothing to the trace (DV projects it to zero). Equivalently, we can choose $V = D^\dagger W$ ($\perp N(D)$ [76]) for any $N \times K$ matrix W, and this change of variables yields a new optimization problem:

$$\langle P \rangle = \operatorname{tr}[(A^{-1}BW)^\dagger O \underbrace{(A^{-1}BW)}_{U}(W^\dagger BW)^{-1}] = \operatorname{tr}\left[U^\dagger O U (W^\dagger BW)^{-1}\right] , \quad (2.22)$$

where $U = A^{-1}BW$ is introduced for simplicity.

Chapter 3

Approaching the upper limits of the local density of states via optimized metallic cavities

In this chapter, we study the special case of a single emitter location with random orientations, which leads to computing the (electric) local density of states (LDOS) at a given location [61]. As discussed in Sec. 2.6, this simplifies the trace expression to a few forward Maxwell solves and conventional methods can be applied to efficiently optimize the LDOS.

First, we give a brief review of the LDOS and related questions (Sec. 3.1), and then briefly review the upper limit of LDOS proposed in a recent work (Sec. 3.2) [95]. Next, we present the methods we used to optimize the LDOS in a metallic cavity (Sec. 3.3). And in Sec. 3.4, we show the optimization results for both total LDOS and single-polarization LDOS, which approach the upper limit within a factor of 10 and 4, respectively. Finally, we give some concluding remarks and discuss possible future improvements (Sec. 3.5).
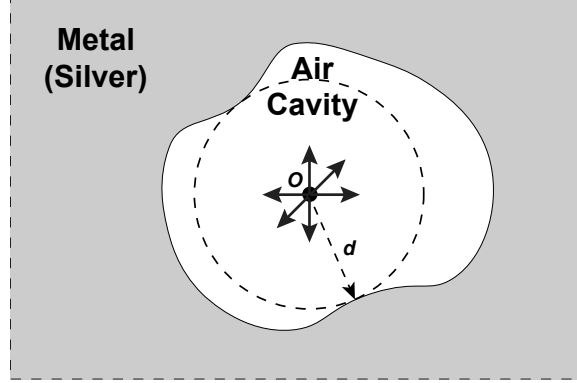
Figure 3-1: Schematic cavity-optimization problem: the shape of an air cavity in a metallic (silver) background is optimized to maximize the LDOS for emitters (dipoles) at the center o, constrained for a minimum separation d (the metal lies outside of a sphere of radius d).

3.1   Introduction

Recently, we obtained theoretical upper bounds [95] to the (electric) LDOS $\rho(\mathbf{x}, \omega)$, a key figure of merit for light–matter interactions (e.g. spontaneous emission) proportional to the power emitted by a dipole current at a position $\mathbf{x}$ and frequency ω [19, 26, 61, 62, 86, 104, 108, 152, 155]. For a resonant cavity with quality factor Q (a dimensionless lifetime), LDOS is proportional to the "Purcell factor" Q/V where V is a modal volume [2, 108], so LDOS is a measure of light localization in space and time. Our LDOS bounds $\sim |\chi|^2 / \operatorname{Im} \chi / d^3$ (reviewed in Sec. 3.2) depend on the material used (described by the ω-dependent susceptibility $\chi = \varepsilon - 1$) and the minimum separation d between the emitter and the material, but are otherwise *independent* of shape, and hence give an upper limit to the localization attainable by *any possible* resonant cavity for (χ, d, ω). However, it is an open question to what extent these bounds are *tight*, i.e. is there any particular cavity design that comes close to the bounds? Initial investigations of a few simple resonant structures were often orders of magnitude from the upper bounds (except at the surface-plasmon wavelength for a given metal) [90, 95, 134]. In this work, we perform computational optimization of 3D metallic cavities at many wavelengths (Fig. 3-1) and find that the bounds are much more nearly attainable than was previously known.

Many previous authors have computationally optimized the LDOS of cavities (or

equivalent quantities such as the Purcell factor Q/V), including many-parameter shape or "topology" optimization [53, 87, 132, 148–150], but in most cases these works did not compare to the recent upper bounds. In many cases, these works studied lossless dielectric materials where the bound diverges (though a finite LDOS is obtained for a finite volume [22, 81] and/or a finite bandwidth [81, 134]). Designs specifically for LDOS of metallic resonators that compared to the bounds initially yielded results far below the bounds except for the special case of a planar surface at the surface-plasmon frequency of the material [93, 95], but recent topology optimization in two dimensions came within a factor of 10 of the 2D bound [22]. Semi-analytical calculations have also been published for resonant modes in spherical metallic voids [63], but did not calculate LDOS. Therefore, the opportunity remains for optimized metallic LDOS designs in three dimensions that approach the theoretical upper bounds. To come as close as possible to the bounds, we focus initially on the idealized case of an air void surrounded by metal filling the rest of space, so that there are no radiation losses; later, we consider the small corrections that arise due to finite metal thickness.

3.2   The local density of states (LDOS)

The (electric) LDOS is equivalent to the total power expended by three orthogonal dipole currents [61]:

$$\rho = \operatorname{Im}\left[\frac{\epsilon_0}{\pi\omega} \sum_{j=1}^{3} \hat{\mathbf{s}}_j \cdot \mathbf{E}_j(\mathbf{x}_0) \right], \tag{3.1}$$

where ϵ_0 is the vacuum electric permittivity, $\mathbf{E}_j$ denotes the field produced by a frequency-ω unit-dipole source at $\mathbf{x}_0$ polarized in the $\hat{\mathbf{s}}_j$ direction, and the sum over j accounts for all three possible dipole orientations. This is equivalent to the average response for *any* dipole orientation [152], and is therefore an isotropic figure of merit. In contrast, we refer to the power $\operatorname{Im}[\frac{\epsilon_0}{\pi\omega}\hat{\mathbf{s}}_j \cdot \mathbf{E}_j(\mathbf{x}_0)]$ expended by only a *single* dipole current as the "polarized" LDOS.

From energy-conservation considerations, previous work found an upper bound for LDOS enhancement inside a cavity compared to vacuum electric LDOS ($\rho_0 =$

$\omega^2/2\pi^2 c^3$ [62]). In particular, the total (electric) LDOS limit can be evaluated as an integral over the entire scattering volume V (the region containing the material χ):

$$\frac{\rho}{\rho_0} \leq 1 + \frac{k^3}{4\pi} \frac{|\chi(\omega)|^2}{\text{Im}\,\chi(\omega)} \int_V \left[\frac{3}{(kr)^6} + \frac{1}{(kr)^4} + \frac{1}{(kr)^2} \right] \mathrm{d}^3 r \,, \tag{3.2}$$

where $\rho_0 = \omega^2/2\pi^2 c^3$ is the free-space electric LDOS [62] and $k = \omega/c$ is the wavenumber.

Ostensibly, this limit is dependent on the exact scattering geometry V (leading to a shape-dependent limit). However, Eq. (3.2) is also an upper bound on any scatterer *contained within* V [95]. In this work, we are interested in a minimal separation d as depicted in Fig. 3-1, so we take V to be a spherical shell with inner radius d and shell thickness L, with $L \to \infty$ for arbitrary thickness. The integral of Eq. (3.2) can then be evaluated as

$$\frac{\rho}{\rho_0} \leq 1 + \frac{|\chi(\omega)|^2}{\text{Im}\,\chi(\omega)} \left[\frac{1}{(kd)^3} + \frac{1}{kd} + \mathrm{O}(kL) \right] \,, \tag{3.3}$$

where $\mathrm{O}(kL)$ is a "Big-O" asymptotic bound [25]. As discussed in Ref. 95, the $\mathrm{O}(kL)$ divergence as $L \to \infty$, which arises from far-field scattering, is unphysical and overly optimistic. The contribution of this term should be limited by the largest interaction distance over which polarization currents contribute to the LDOS, thus is generally small compared to the first two terms on the right-hand side of Eq. (3.3) and can be neglected at small separation distance d. Therefore, the total LDOS limit for a metallic cavity with minimum separation distance d is

$$\frac{\rho}{\rho_0} \leq 1 + \frac{|\chi(\omega)|^2}{\text{Im}\,\chi(\omega)} \left[\frac{1}{(kd)^3} + \frac{1}{kd} \right] \,. \tag{3.4}$$

Note that this limit, where V is the exterior of a sphere, is about 8 times larger than the limit discussed in Ref. 95 where V was a planar half-space. In practice, this factor-of-8 improvement may be difficult to realize since optimized cavities will typically have only a small surface area at the minimum separation d (except for the resonant spheres discussed in Sec. 3.4 below).

For the polarized LDOS limit, the integral in Eq. (3.2) (squared Frobenius norm

of the homogeneous Green's function [95]) is replaced with the norm of the dipole polarization vector multiplied by the Green's function norm [90]:

$$\frac{\rho_\text{p}}{\rho_0} \leq \frac{1}{3} + \frac{k^3}{8\pi} \frac{|\chi(\omega)|^2}{\text{Im}\,\chi(\omega)} \int_V \left[a(r) + b(r)|\hat{\mathbf{p}} \cdot \hat{\mathbf{r}}|^2\right] \text{d}^3 r \;, \tag{3.5}$$

where $\hat{\mathbf{p}}$ is the unit vector in the polarization direction, and $a(r)$ and $b(r)$ are:

$$a(r) = \frac{1}{(kr)^6} - \frac{1}{(kr)^4} + \frac{1}{(kr)^2} \tag{3.6}$$

$$b(r) = \frac{3}{(kr)^6} + \frac{5}{(kr)^4} - \frac{1}{(kr)^2} \;. \tag{3.7}$$

Similar to the total LDOS limit analysis, we can use a spherical bounding surface of radius d to derive a general upper bound (also excluding the diverging $\text{O}(kL)$ term):

$$\frac{\rho_\text{p}}{\rho_0} \leq \frac{1}{3} + \frac{|\chi(\omega)|^2}{3\,\text{Im}\,\chi(\omega)} \left[\frac{1}{(kd)^3} + \frac{1}{kd}\right] \;, \tag{3.8}$$

which is exactly $1/3$ of the total LDOS limit. That is, the total LDOS bound is equivalent to assuming that the polarized LDOS bound can be attained for all three polarizations simultaneously, which our results show to be unlikely.

To compute the shape-dependent polarized-LDOS bound (for a given optimized shape) in Sec. 3.4.3, we performed numerical integration of Eq. (3.5) over spherical angles (with the r integral performed analytically), but excluding the $1/(kr)^2$ radiative term that yields the $\text{O}(kL)$ divergence.

It is important to emphasize that the derivation of Eq. (3.4) and Eq. (3.8) gives a rigorous upper bound to the LDOS, but does not say what structure (if any) achieves the bound. By actually solving Maxwell's equations for various geometries, we can investigate how closely the bound can be approached (how "tight" the bound is). It is possible that incorporating additional constraints may lead to tighter bounds in the future [40,71,97,99], but our results below already show that Eq. (3.4) and Eq. (3.8) are achievable within an order of magnitude.

3.3 Cavity-optimization methods

To numerically compute the LDOS inside a metal cavity, we employed a free-software implementation [122] of the boundary element method (BEM) [7]. A BEM formulation only involves unknown tangential fields on the metal surface, leading to modest-size computations for 3D metallic voids (Fig. 3-1). The complex dielectric constant of silver was interpolated from tabulated data [110]. In addition, we implemented an adjoint method [98, 138] to rapidly obtain the gradient of the LDOS with respect to the shape parameters described below.

The gradient of the LDOS with respect to many shape parameters can be computed by solving Maxwell's equations a *single* additional time (for "adjoint" fields) via the adjoint method, a key algorithm for large-scale photonics optimization [98, 138]. The specific case of a boundary perturbation is reviewed in Ref. 92, which shows that the variation of an objective function F in response to small shape deformations δR (the surface displacement in the normal direction) over the surface $\partial \Omega$ is

$$\delta F = 2 \operatorname{Re} \iint_{\partial \Omega} \delta R(\mathbf{x}') \left[(1 - \varepsilon) \mathbf{E}_{\parallel}(\mathbf{x}') \cdot \mathbf{E}_{\parallel}^{\mathrm{A}}(\mathbf{x}') + \left(\frac{1}{\varepsilon} - 1 \right) \mathbf{D}_{\perp}(\mathbf{x}') \cdot \mathbf{D}_{\perp}^{\mathrm{A}}(\mathbf{x}') \right] \mathrm{d}S,$$
$$(3.9)$$

where ε is the electric permittivity of the metal, $\mathbf{E}_{\parallel}$ is the surface-parallel electric field, $\mathbf{D}_{\perp}$ is the surface-parallel displacement field, and the superscript "A" denotes the adjoint field excited by an adjoint current source $\mathbf{J} = \partial F / \partial \mathbf{E}$.

In the case of LDOS, a further simplification arises. The objective function $F = \rho$ at position $\mathbf{x}_0$ can be expressed as Eq. (3.1). We can see from Eq. (3.1) that the LDOS is proportional to the electric field, leading to an adjoint field that is also proportional to the original problem for each orientation j,

$$\mathbf{E}_j^{\mathrm{A}}(\mathbf{x}) = \operatorname{Im} \left[\frac{\epsilon_0}{\pi \omega} \mathbf{E}_j \right] . \qquad (3.10)$$

Inserting Eq. (3.10) into Eq. (3.9) gives us the LDOS gradient (first-order variation)

with respect to any shape deformation:

$$\delta\rho = \frac{\epsilon_0}{\pi\omega}\,\mathrm{Im}\sum_j \iint_{\partial\Omega}\delta R(\mathbf{x}')\left[(1-\varepsilon)\mathbf{E}_{j\parallel}(\mathbf{x}')^2 + \left(\frac{1}{\varepsilon}-1\right)\mathbf{D}_{j\perp}(\mathbf{x}')^2\right]\mathrm{d}S\,. \qquad (3.11)$$

We can see that the gradient of LDOS with respect to all shape parameters simultaneously is obtained by the adjoint method using only two BEM simulations—the original problem and an adjoint problem (the same Maxwell problem with artificial "adjoint" sources). Since the adjoint problem is the same Maxwell/BEM operator, we need only form and factorize the BEM matrix a single time, and the computational cost to solve both the forward and adjoint problems is essentially equivalent to a single simulation. Validation against a semi-analytical solution for spheres [63] is discussed in Sec. 3.4.2.

In order to parameterize an arbitrary cavity shape numerically, we use a level-set description [106,146], combined with a free-software surface-mesh generator CGAL [15, 141]. In particular, we describe the radius of the cavity around the source point by a function $R(\theta,\phi)$ (in spherical coordinates), which is expanded below in either spherical harmonics or other polynomials, and equivalently pass a level-set function $\Phi = r - R(\theta,\phi)$ to CGAL (such that $\Phi = 0$ defines the surface). We considered various parameterizations of the shape function R. The simplest geometries considered were ellipsoids, cylinders, or rectangular boxes, described by two or three parameters. For many-parameter optimization with a minimum radius (separation) d, the function R is expressed as an expansion in some basis functions $S_n(\theta,\phi)$ as:

$$R(\theta,\phi) = d + \left|\sum_{n=0}^{N}c_n S_n(\theta,\phi)\right|^2. \qquad (3.12)$$

For the basis functions S_n, we used either spherical harmonics $Y_{\ell m}(\theta,\phi)$ (for arbitrary asymmetrical "star-shaped" cavities) or simple polynomials in θ (to impose azimuthal

symmetry round the z axis and a $z = 0$ mirror plane):

$$S_n(\theta, \phi) = \begin{cases} \theta^n & \theta \leq \frac{\pi}{2}, \\ (\pi - \theta)^n & \theta > \frac{\pi}{2}. \end{cases} \tag{3.13}$$

The level set is discretized for BEM (by the CGAL software) into a triangular surface mesh. We used $5\times$ greater resolution for surface points closer to the dipole source (radius $\lesssim 70$ nm), since the singularity of the fields at the source point leads to rapid variations nearby, for around 5000 triangles overall. As we deformed the shape during optimization, we first deform the triangles smoothly as long as all angles remained between 30° and 120°, after which point we triggered a re-meshing step. Unfortunately, re-meshing causes slight discontinuities in the objective function and its derivatives which tend to confuse optimization algorithms expecting completely smooth functions [142]. We tried various optimization algorithms designed to be robust to such "numerical noise" [46, 65], and found that the Adam stochastic-optimization algorithm [65] seems to work best for our problem.

3.4 Cavity-optimization results

3.4.1 Total LDOS

We performed numerical shape optimization of the LDOS for cavities formed by voids in silver [110] at wavelengths λ from 400–900 nm, for both simple geometries (cylinder, ellipsoid, and rectangular box) and complex many-parameter shape (spherical harmonics). To obtain a finite optimum LDOS, one must choose a lower bound on the emitter–metal separation distance d [95]. We chose $d = 50$ nm so that $kd < 1$ for all optimized wavelengths; this allows us to use Eq. (3.4) as the LDOS upper bound, neglecting additional far-field effects [95], while a much smaller d was inconvenient to model (due to extremely small feature sizes and even nonlocal effects [121] at such scales). Note that d also sets a lengthscale for the region of the cavity with maximum LDOS: as long as one shifts the emitter location by $\ll d$ ($\lesssim 10$ nm), the optimized
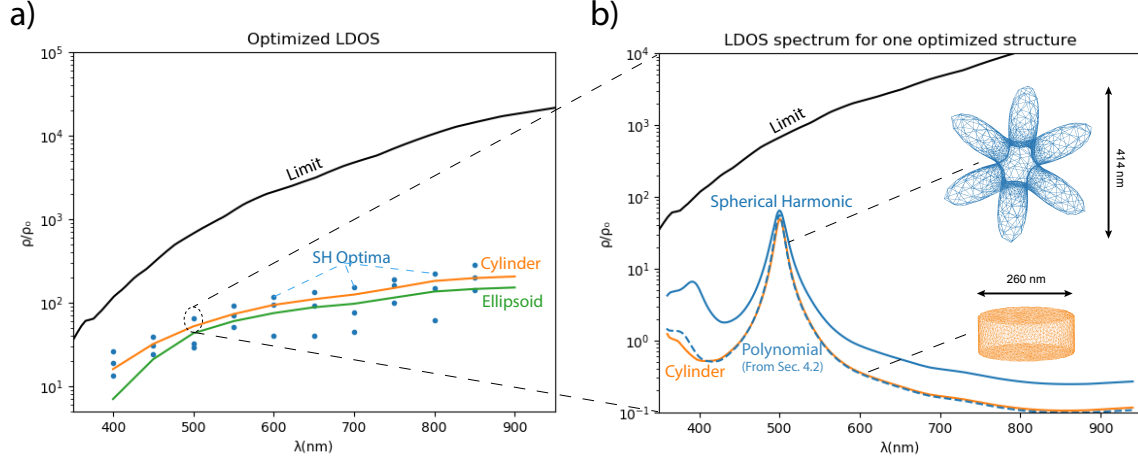
Figure 3-2: (a) Total LDOS optima as a function of the wavelength λ for a minimum separation $d = 50$ nm, along with the upper bound (black line). A separately optimized structure is used for each wavelength, either optimized cylinders (orange line) and ellipsoids (green line) or general shape optimization via the optimized spherical-harmonic (SH) surfaces (blue dots) of Eq. (3.12). Several SH local optima are shown for each λ, whereas for cylinders and ellipsoids only the global optima are shown. (b) LDOS spectra of the spherical-harmonic (blue) and cylinder (orange) structures optimized for $\lambda = 500$ nm, the the shapes (not to scale) inset (see supplementary Visualizations 1 and 2 for 3D views). Also shown is the total-LDOS spectrum of a polynomial shape (dashed blue line) optimized for the polarized LDOS in Sec. 3.4.3, showing that optimizing for a single dipole orientation (polarized LDOS) is nearly equivalent in performance to optimizing for all orientations (total LDOS).

LDOS will be of the similar magnitude, but other (unoptimized) locations in the cavity will typically have a drastically different LDOS.

The results of the optimized LDOS as a function of the wavelength are displayed in Fig. 3-2a. Note that each wavelength corresponds to a *different* structure optimized for that particular wavelength. If we fix the structure as the one optimized for $\lambda = 500$ nm, the resulting LDOS spectrum is shown in Fig. 3-2b, exhibiting a peak at the optimized wavelength. For simple shapes (cylinder and ellipsoid), we swept the parameters through a large range and found the global optimum among several local optima. We also optimized rectangular boxes, but their performance was nearly identical to that of the cylinders (but slightly worse), so they are not shown. For the 16-parameter (spherical-harmonic) level-set optimization, we performed local optimization for ~ 10 random starting points and plotted the best result along with

a few other typical local optima.

We found that the optimized LDOS comes within a factor of 10 of the upper bound in the short-wavelength regions ($\lambda < 550$ nm), and the optimized cylinders are surprisingly good (within $\approx 20\%$ of the many-parameter optima). The optimized cavity geometries at $\lambda = 500$ nm are shown in the inset of Fig. 3-2b. We can see that the optimized many-parameter shape has a three-fold rotational symmetry around one axis; consequently, it has equal polarized LDOS in two directions but $\sim 100\times$ smaller polarized LDOS in the third direction. The spherical-harmonic basis is unitarily invariant under rotations, so this means that the optimization of the total LDOS (an isotropic figure of merit) exhibits a spontaneous symmetry breaking: it chooses two directions to improve at the expense of the third. For the optimized cylinder and ellipsoid, the polarized LDOS is only large for *one* polarization (along the cylinder axis, the "short" axis). A similar spontaneous symmetry breaking was observed for optimization of LDOS in two dimensions [81], as well as in saturating the upper bounds for scattering and absorption [91, 95] (where it was related to quasi-static sum rules constraining polarizability resonances [91]).

As discussed below, we found that we could approach the polarized LDOS bound at $\lambda = 500$ nm within a factor of ~ 3 for a single dipole orientation; the fact that the total LDOS optimization is worse compared to its bounds ($\approx 3 \times$ polarized bound) reiterates the conclusion that it is probably not generally possible to maximize the polarized LDOS for all three directions simultaneously.

One possible shape with *uniform* polarized LDOS in all directions is a sphere. As a matter of fact, we found that at each wavelength $\lesssim 600$ nm, there exists a resonant sphere [14, 63] such that the LDOS at the center of the spheres comes within $\approx 20\%$ of the corresponding-d bound . However, this resonant radius d is relatively large ($\lambda/4 < d < \lambda/2$, in order to create a resonance at λ) leading to small LDOS and bound ($\rho/\rho_0 \sim 10$–100), so saturating such a large-d bound may have limited utility. Spheres at much smaller d do not exhibit these "void resonances" and have much worse LDOS than the asymmetrical shapes in Fig. 3-2 for $d = 50$ nm.
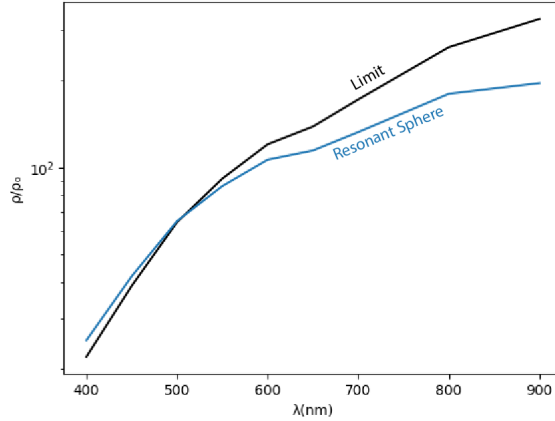
Figure 3-3: LDOS of a resonant air sphere in silver as a function of the wavelength λ (blue line), where for each λ we choose the smallest radius a_{res} for which we couple to a resonant mode at λ. The black line is the corresponding upper bound from Eq. (3.4), setting the minimum separation distance $d = a_{\text{res}}$. The LDOS slightly exceeds the bound at small wavelengths where the radius becomes so large that one would need to include the $O(kL)$ term that we dropped in Eq. (3.3).

3.4.2 Resonant Sphere

For a void sphere cavity, the resonant electromagnetic surface modes can be analytically obtained by solving the equation [14,63] (after correcting a typographical error in Ref. 63):

$$\epsilon_{\text{m}}(\omega)H_\ell(k_{\text{m}}a)\left[k_{\text{d}}aJ_\ell(k_{\text{d}}a)\right]' = \epsilon_{\text{d}}J_\ell(k_{\text{d}}a)\left[k_{\text{m}}aH_\ell(k_{\text{m}}a)\right]' \ , \tag{3.14}$$

where a corresponds to the void radius, ℓ is the (integer) index denoting the angular momentum, $k_{\text{m}} = \sqrt{\epsilon_{\text{m}}}k$ and $k_{\text{d}} = \sqrt{\epsilon_{\text{d}}}k$ are wave vectors in metal and air/vacuum respectively, J_ℓ and H_ℓ are spherical Bessel and Hankel functions of the first kind, and the prime denotes differentiation with respect to $k_{\text{d}}a$ or $k_{\text{m}}a$. Since the excitation source in our LDOS problem is a dipole at the center of the sphere, only an $\ell = 1$ mode can be excited. Therefore, for each wavelength $2\pi/k$, there is a minimal resonant sphere: a minimal radius $d = a_{\text{res}}$ satisfying Eq. (3.14) for $\ell = 1$.

Using Eq. (3.14), we can directly compute this minimal resonance radius a_{res} and then evaluate the corresponding LDOS. Furthermore, since a semi-analytical solution for the resonant modes is known for a void sphere [63], the LDOS can also

be evaluated analytically via Eq. (3.1). We find that our BEM results are within 2% of the analytical values using a surface mesh of ≈ 5000 triangles, validating our numerical solver. The resulting LDOS values are shown in Fig. 3-3, which shows that the LDOS of the resonant sphere is very close to the theoretical limit (within $\approx 10\%$) for the corresponding minimal separation $d = a_{\text{res}}$. These strong results verify the limit at least at the resonance combinations of d and λ.

Notice that the resonant-sphere LDOS seems to actually slightly exceed the theoretical limit obtained with Eq. (3.4) in short wavelength range. There is no contradiction however: this is simply the effect of the $O(kL)$ we dropped in Eq. (3.3). In particular, the resonant radius here is relatively large compared to the wavelength. For example, at $\lambda = 700$ nm we get $a_{\text{res}} = 273$ nm, for which $kd = 2.45$ and $1/(kd)^3 + 1/(kd) = 0.47$. As a matter of fact, $1/(kd)^3 + 1/(kd) < 1$ for all resonant spheres, thus the $O(kL)$ term in Eq. (3.3) will have a non-negligible influence on the bound, causing the actual bound to be slightly higher than Eq. (3.4).

3.4.3 Polarized LDOS

Since the spontaneous symmetry-breaking in the previous section suggests that optimization favors maximizing LDOS in a single direction, we now consider optimizing the polarized LDOS. That is, we maximize the power expended by a dipole current with a single orientation (similar to previous work on cavity optimization in dielectric media [22, 81, 150]). As above, we performed few-parameter optimization of ellipsoids, cylinders, and rectangular boxes. For many-parameter optimization, we initially used spherical harmonics but observed that optimizing polarized LDOS naturally leads to structures that are rotationally symmetric around the dipole axis. To exploit this fact, we switched to simple polynomials in θ as described in Sec. 3.3. Specifically, we first performed a rough scan of degree-2 polynomials to obtain a starting point, then we performed a degree-5 optimization using the adjoint method (degree-10 gave similar results at greater expense). (Gradually increasing the number of degrees of freedom is "successive refinement," a heuristic that has also been used in other work to avoid poor local minima [109, 135].) The results are shown in Fig. 3-4. We only
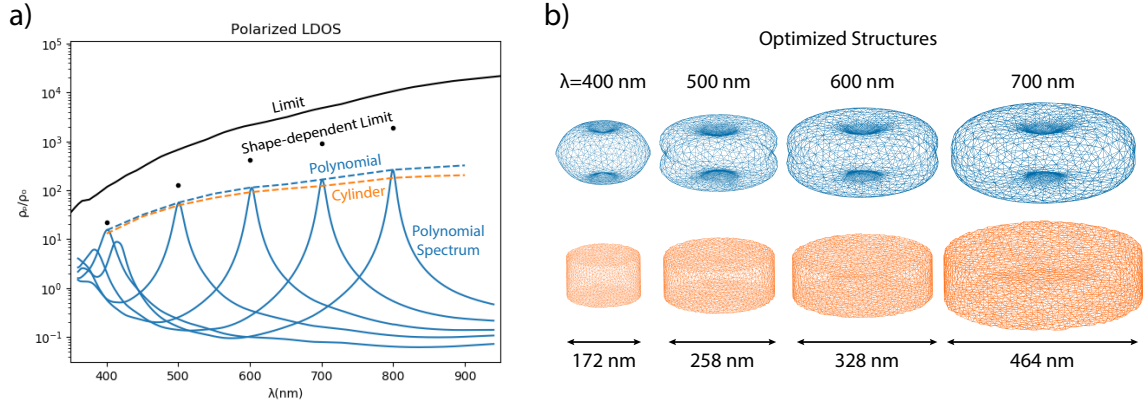
Figure 3-4: (a) Polarized LDOS optima (dashed lines) as a function of the wavelength λ at a minimum separation $d = 50$ nm, along with the upper bound (black line) and the shape-dependent bounds (black dots). Dashed lines are peak performance of separately optimized structures for each λ, either cylinders (orange) or the optimized polynomials (blue) of Eq. (3.13). Solid blue lines are spectra of the polarized LDOS for optimized polynomial designs at selected wavelengths $\lambda = 400, 500, 600, 700, 800$ nm, respectively. (b) Optimized polynomial (top) and cylinder (bottom) structures at the wavelength $\lambda = 400, 500, 600, 700$ nm (to scale; see supplementary Visualizations 3–10 for 3D views).

plotted the cylinder results (orange dashed line), because the ellipsoid and box results were worse.

We obtain an optimized LDOS within a factor of about 4 of the polarized-LDOS bound in the short wavelength regions ($\lambda < 550$ nm). At a wavelength of 400 nm, the optimized LDOS is only 2.5 times smaller than the bound, which greatly improves upon previous results that often came only within 10^2–10^3 of the bound [93, 95, 134]. One interesting fact is that the optimized polarized LDOS is actually only slightly smaller ($\approx 10\%$) than the optimized total LDOS (blue dashed line in Fig. 3-2b), which is consistent with the spontaneous symmetry breaking we commented on above: optimizing total LDOS spontaneously chooses one or two directions to optimize at the expense of all others, and hence is often equivalent to optimizing polarized LDOS. (If an isotropic LDOS is required by an application, one approach is to maximize the minimum of three polarized LDOSes [81].)

The upper bounds can help us to answer another important question: how much additional improvement could be obtained by introducing additional void structures

outside of our cavity? (For example, by giving the cavity walls a finite thickness.) An upper bound to this improvement is provided by computing a *shape-dependent limit*: we use the same bounding procedure, but evaluate the limit assuming the material lies *outside our optimized shape* rather than outside of a bounding sphere. This analysis, which is carried out in Sec. 3.2, shows that our optimized polarized LDOS is nearly reaching this shape-dependent limit as shown by the black dots in Fig. 3-4a. Therefore, little further improvement is possible using additional structures outside of the cavity, which justifies optimizing over simple voids in order to probe the bounds.

3.4.4 Radiative LDOS from a finite thickness cavity

In the previous sections, we optimized the *total* and *polarized non-radiative* LDOS for an air-void cavity as in Fig. 3-1 (since the power is entirely absorbed). However, the same theoretical procedure yielded a bound on the purely *radiative* power that was simply 1/4 of the total-LDOS bound [95], and in fact the two results are closely related. In general, the addition of low-loss input/output channels can be analyzed via coupled-mode theory as small perturbations to existing cavity designs [55].

Given a purely absorbing resonant cavity such as the ones optimized in the previous sections, the addition of a radiation-loss channel (e.g. by coupling the cavity to a waveguide, permitting radiation through a small hole in the cavity walls, or simply thinning a portion of the wall) can be analyzed quantitatively using the technique of temporal coupled-mode theory (TCMT) as long as the lifetime of the cavity remains long enough to be treated as an isolated resonant mode [55, 139]. TCMT yields a result identical to that of Ref. 95: the optimal radiated power is 1/4 of the original absorbed power. We describe that straightforward analysis here, because it helps to connect the results in the previous sections with applications to radiative cavities.

In particular, consider a purely absorptive cavity with a quality factor [55, 139] $Q_a \gg 1$, so that the Purcell enhancement of the LDOS for a dipole source is proportional to Q_a/V where V is a corresponding modal volume [55,108]. Now, suppose that one perturbs the cavity to add a radiative loss channel with a corresponding quality

factor Q_r. As long as the radiation loss is low ($Q_r \gg 1$, as is necessary to retain a well-defined resonance), it can be treated as a small perturbation and the effect on Q_a and V can be neglected as higher-order in $1/Q_r$ [55]. With the addition of this channel, the total cavity Q becomes $Q = Q_a Q_r/(Q_a + Q_r)$, corresponding to a total nondimensionalized loss rate of $1/Q = 1/Q_a + 1/Q_r$ [55], so the Purcell enhancement factor is reduced to Q/V. However, only a fraction Q/Q_r of the dipole power goes into radiation (vs. absorption), so the resonant enhancement of the radiated power is proportional to

$$\frac{Q^2}{VQ_r} = \frac{Q_a^2 Q_r}{(Q_a + Q_r)^2 V}. \tag{3.15}$$

Equation (3.15) is maximized when $Q_r = Q_a$, i.e. when the absorptive and radiative loss rates are *matched*, similar to results obtained previously [42,131]. For $Q_r = Q_a$, Eq. (3.15) becomes $Q_a/4V$, or exactly $1/4$ of the Q_a/V Purcell enhancement in the purely radiative case.

This result is consistent with the fact that our radiative-LDOS bound is $1/4$ of the total LDOS bound [95], but is more far-reaching. It prescribes how any high-Q absorbing cavity can be converted to a radiating cavity with about $1/4$ the radiated power, similar to the results we obtained numerically below.

To conclude, given an resonant absorptive cavity, one could modify it to radiate at most $\approx 1/4$ of the original LDOS by slightly perturbing it to add a radiative-escape channel (e.g. a small hole or thinning in the cavity wall) tuned to *match* the absorption-loss rate [42,131]. (The same channel could also be used to introduce *input* energy, e.g. for pumping an emitter.) In the following we will show an example of this: by thinning the cavity walls, we achieve radiated power $\approx 1/4$ of the absorbed power in the purely absorbing cavity.

Here, we study the effect of a finite thickness of the metallic walls, replacing the infinite metallic regions of Fig. 3-1. To do this, we took the optimized cylinder at $\lambda = 500$ nm from Fig. 3-2b and modified the silver walls to have finite thickness with the same inner surface. The LDOS as a function of the shell thickness is shown in Fig. 3-5. We observe that the LDOS increases monotonically with the shell thickness,
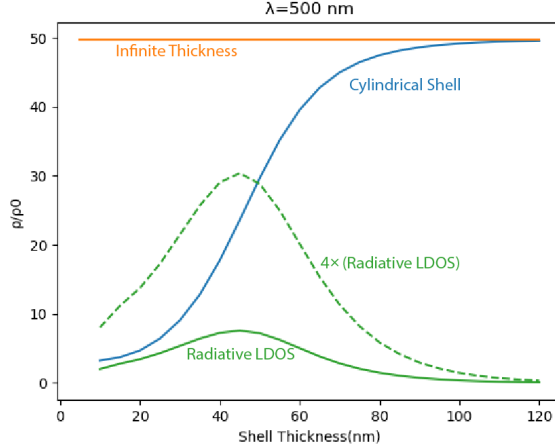
Figure 3-5: LDOS of optimized cylindrical cavity a function of the wall thickness (blue line) at a wavelength $\lambda = 500$ nm and a minimum separation $d = 50$ nm, compared to the infinite-thickness LDOS (orange line). Also shown are the *radiative* LDOS $\rho_{\rm rad}/\rho_0$ (solid green line: the radiated/non-absorbed power), as well as $4 \times \rho_{\rm rad}/\rho_0$ (dashed green line) because the theoretical bounds predict that the maximum $\rho_{\rm rad}$ is $1/4$ of the total LDOS [95].

and that a shell thickness of about 100 nm (about 3.7 times the skin depth) yields a polarized LDOS within 5% of the infinite-thickness result, which is not surprising considering that the skin depth [51, 110] of silver is < 33 nm for $\lambda > 400$ nm.

For a finite-thickness shell, some of the expended power (total LDOS) is absorbed and some "leaks" through the finite thickness to radiate away, and it is interesting to consider the *radiative* LDOS $\rho_{\rm rad}$ defined as the latter radiated power for the same dipole source (green line in Fig. 3-5). As a function of shell thickness, $\rho_{\rm rad}$ exhibits a peak: too thin and the resonance is too weak to enhance LDOS, but too thick and no power escapes to radiate (all power is absorbed). The theoretical limit for $\rho_{\rm rad}/\rho_0 - 1$ is $1/4$ of the limit for the total (absorbed+radiated) LDOS [95]. Correspondingly, we plot $4\rho_{\rm rad}/\rho_0$ in Fig. 3-5 (dashed green line) and see that, at the optimum $\rho_{\rm rad}$, the radiative LDOS is approximately $1/4$ of the total ($4\rho_{\rm rad} \approx \rho$), agreeing with the prediction of polarization-maximization in Ref. 95.

3.5   Concluding remarks

In this work, we obtain LDOS values within a factor of ≈ 10 of the total LDOS bound and a factor of ≈ 4 of the polarized LDOS bound in a many-parameter metal-cavity optimization, showing that these upper bounds are much more nearly attainable than was previously known [95]. Unlike previous work on scattering/absorption by small particles [91], our optimized cavities do not appear to be in the quasi-static regime, since their largest axes are $\sim \lambda/2$, even while their smallest axes are deeply subwavelength and exhibit strong plasmonic effects.

It is possible that further improvements could be obtained by a more extensive search of local optima, or by expanding the search to other classes of cavities beyond "star-shaped" structures that can be described by a $R(\theta, \phi)$ level set, e.g. via full 3D topology optimization. We would also like to systematically explore the attainability of the finite-bandwidth bounds from Ref. 134, which are useful for lossless dielectrics (where the single-frequency LDOS bound diverges). Conversely, it is possible that incorporating additional constraints, such as considering a more complete form of the optical theorem, may lower the LDOS bounds [40, 71, 97, 99]. There has also been recent interest in the magnetic LDOS, corresponding to magnetic-dipole radiation [9], and we expect that qualitatively similar results (albeit with different optimal shapes) would be obtained for the magnetic LDOS (or some combination of magnetic and electric, although trying to optimize both simultaneously would likely encounter difficulties similar to those for optimizing multiple polarizations). The magnetic LDOS would merely require one to replace our electric-dipole source with a magnetic-dipole source, along with a similar switch of the Green's function that appears in the derivation of the bounds [95].

On a more practical level, a possible next step is to maximize LDOS (or similar figures of merit) for 3D geometries more amenable to fabrication, whereas our goal in the present paper was to probe the fundamental LDOS limits without concern for fabrication. Fortunately, our results show that relatively simple (constant cross-section) shapes such as cylinders can perform nearly as well as the irregular shapes

produced by many-parameter shape optimization, and are relatively insensitive to small details (e.g. curved or flat walls). This is a hopeful sign for adapting such cavities to nano-manufacturing by lithography or other techniques. And, although infinite-thickness cavities completely absorb the emitted power, our computation of the radiated power in finite-thickness shells (Section 3.4.4) agrees with the theoretical bound's prediction that the optimal radiated power is $\approx 1/4$ of the total [95].

Chapter 4

Topology optimization and its application for incoherent emission

In this chapter, we first briefly review the key ideas in density-based topology optimization (TopOpt, Sec. 4.1) and the detailed numerical formulation for a 2D electromagnetic scheme (Sec. 4.2). We then illustrate our method with several 2D example problems (Sec. 4.3: fluorescence from an optimized nanoparticle, enhanced emission from a corrugated surface analogous to a light-emitting diode [32, 52], and optimized emission into a waveguide. Finally we discuss the potential application of the trace formulation and possible improvements in future works (Sec. 4.4).

Density-based TopOpt has attracted increasing interest over the last few decades because of its ability to reveal surprising high-efficiency designs by optimizing over thousands or even millions of design degrees of freedom [53]. It parameterizes a structure by an artificial "density" $\rho(\mathbf{x}) \in [0, 1]$ at *every point* (or every "pixel") in a design region, which is typically passed through smoothing and threshold steps to yield a physical "binary" design consisting of one of two materials at every point. We apply a damped-diffusion filter [77], which regularizes the problem by setting a minimum lengthscale on the design. (Additional manufacturing constraints can be imposed by well-known techniques [43], but in the present work we focus on the fundamental algorithms and not on experimental realization.) Once a scalar objective function (to be optimized) is defined, such as the emitted power (e.g. the new for-

mulation in this paper), its derivatives (sensitivities) with respect to all the design parameters can be efficiently computed with a single additional simulation via adjoint methods [98, 142]. Given the objective function and its derivatives, a variety of large-scale optimization algorithms are available; we use the CCSA/MMA method [140]. We employ a recent free/open-source finite-element method (FEM) package, Gridap.jl [6], in the Julia language [13], which allows us to efficiently code highly customized FEM-based trace formulations in a high-level language, with the construction of the adjoint problem aided by automatic-differentiation (AD) tools [49, 125]. In `https://gridap.github.io/Tutorials/dev/pages/t018_TopOptEMFocus/`, we wrote a tutorial about how to use Gridap.jl for TopOpt.

4.1   Topology-optimization formulation

In this section, we briefly review the density-based TopOpt formulation [53] that we employ for our example applications in Sec. 4.3. The key idea of TopOpt is that an "artificial density" field, $\rho(\mathbf{x}) \in [0, 1]$, is defined on a spatial "design" domain. This field is then filtered (to impose a non-strict minimum length-scale) and thresholded (to mostly "binarize" the geometry, resulting in a physically admissible geometry). The resulting smoothed and thresholded field is then used to control the spatial material distribution, constituting the structure under design. The design field, ρ, is discretized into a finite number of design degrees of freedom, which constitutes the design variables in the inverse design problem to be solved, e.g. Eq. (2.20), using a finite-element method (FEM) on a triangular mesh [6, 54], and the geometry is optimized using a well-known gradient-based algorithm that scales to high-dimensional problems with thousands or millions of degrees of freedom [140].

Given a density $\rho(\mathbf{x}) \in [0, 1]$, one should first regularize the optimization problem by setting a non-strict minimum lengthscale r_f, as otherwise one may obtain arbitrarily fine features as the spatial resolution is increased. This is achieved by convolving ρ with a low-pass filter to obtain a smoothed density $\tilde{\rho}$ [53]. There are many possible filtering algorithms, but in an FEM setting (with complicated nonuniform meshes), it

is convenient to perform the smoothing by solving a simple "damped diffusion" PDE, also called a Helmholtz filter [77]:

$$-r_f^2 \nabla^2 \tilde{\rho} + \tilde{\rho} = \rho \,,$$
$$\left. \frac{\partial \tilde{\rho}}{\partial \mathbf{n}} \right|_{\partial \Omega_D} = 0 \,, \tag{4.1}$$

where r_f is the lengthscale design parameter and $\mathbf{n}$ is the normal vector at the boundary $\partial \Omega_D$ of the design domain Ω_D. This damped-diffusion filter essentially makes $\tilde{\rho}$ a weighted average of ρ over a radius of roughly r_f [77]. (In addition to this filtering, it is possible to impose additional fabrication/lengthscale constraints, for example to comply with semiconductor-foundry design rules [43].)

Next, one employs a smooth threshold projection on the intermediate variable $\tilde{\rho}$ to obtain a "binarized" density parameter $\tilde{\tilde{\rho}}$ that tends towards values of 0 or 1 almost everywhere [151]:

$$\tilde{\tilde{\rho}} = \frac{\tanh(\beta\eta) + \tanh\left(\beta(\tilde{\rho} - \eta)\right)}{\tanh(\beta\eta) + \tanh\left(\beta(1 - \eta)\right)} \,, \tag{4.2}$$

where β is a steepness parameter and $\eta = 0.5$ is the threshold. During optimization, one begins with a small value of β (allowing smoothly varying structures) and then gradually increases β to progressively binarize the structure [22]; in the examples below, we used $\beta = 5, 10, 20, 40, 80$, similar to previous authors.

Finally, one obtains a material, described by an electric relative permittivity (dielectric constant) $\varepsilon(\mathbf{r})$ in Eq. (2.1), given by:

$$\varepsilon(\mathbf{r}) = \left[\varepsilon_1 + (\varepsilon_2 - \varepsilon_1)\tilde{\tilde{\rho}}(\mathbf{r})\right] \left(1 + \frac{\mathrm{i}}{2Q}\right) \,, \tag{4.3}$$

where ε_1 is the background material (usually air, $\varepsilon_1 = 1$) and ε_2 is the design material (we use dielectric of $\varepsilon_2 = 12$ in this chapter).

Equation (4.3) includes an optional "artificial loss" term $\sim 1/Q$, which effectively smooths out resonances to have quality factors $\leq Q$ (fractional bandwidth $\geq 1/Q$) [81]. Such an artificial loss is useful in single-ω emission optimization in order to set a minimum bandwidth of enhanced emission, rather than obtaining diverging

enhancement over an arbitrarily narrow bandwidth as is possible with lossless dielectric materials [81]. Also, optimizing low-Q resonances often leads to better-behaved optimization problems (less "stiff" problems with faster convergence), so during optimization we start with a low $Q = 5$ and geometrically increase it (to $Q = 1000$) as the optimization progresses [81]

Numerically, we use finite element method to discretize the computation domain to standard triangular meshes with first-order Lagrange elements [54] and impose perfectly matched layers (PMLs) for absorbing boundaries [107]. We discretized ρ and $\{\tilde{\rho}, \tilde{\tilde{\rho}}\}$ with piecewise-constant (0th-order) and first-order elements, respectively. During optimization, one must ultimately compute the sensitivity of the objective function (the trace from Sec. 2.4) with respect to the degrees of freedom ρ—for each step outlined above (smoothing, threshold, PDE solve, etcetera) we formulate a vector–Jacobian product following the adjoint method for sensitivity analysis [98,142] with some help from automation [125], and then these are automatically composed ("backpropagated") by an automatic-differentiation (AD) system [49]. In this way, the gradient with respect to all of the degrees of freedom (ρ at every mesh element) can be computed with about the same cost as that of evaluating the objective function once [98]. More details are presented in Sec. 4.2.

4.2   Numerical formulation

In this section, we provide details of the mathematical formulation and numerical implementation that we used for the examples in Sec. 4.3.

We employ the frequency-domain Maxwell equations for the magnetic field $\mathbf{H}$ arising from an electric current $\mathbf{J}$ with a dielectric function (relative permittivity) ε and a relative magnetic permeability μ:

$$\left[\nabla \times \frac{1}{\varepsilon} \nabla \times - \left(\frac{\omega}{c} \right)^2 \mu \right] \mathbf{H}(\mathbf{x}) = \nabla \times \left[\frac{1}{\varepsilon} \mathbf{J}(\mathbf{x}) \right]. \tag{4.4}$$

For 2D (z-invariant) problems, we chose in-plane currents $\mathbf{J}$, so that the resulting

magnetic fields $\mathbf{H} = H_z\hat{\mathbf{z}}$ are polarized purely in the z direction [55]. In this case Eq. (4.4) simplifies to a scalar Helmholtz equation:

$$\left[-\nabla \cdot \frac{1}{\varepsilon}\nabla - \left(\frac{\omega}{c}\right)^2 \mu\right] H_z = \left(\nabla \times \left[\frac{1}{\varepsilon}\mathbf{J}(\mathbf{x})\right]\right) \cdot \hat{\mathbf{z}}. \tag{4.5}$$

Note that, for the correlation functions in the previous discussion, we simplified the right-hand side by absorbing the $1/\varepsilon$ scaling into $\mathbf{J}$.

We employ perfectly matched layers (PMLs) for absorbing boundaries, with Dirichlet ($u = 0$) boundary conditions behind the PML. The implementation of the "stretched-coordinate" PML is simply a replacement $\nabla \to \Lambda\nabla$ in Eq. (4.5) [54, 107]:

$$\left[-\Lambda\nabla \cdot \frac{1}{\varepsilon}\Lambda\nabla - \left(\frac{\omega}{c}\right)^2 \mu\right] H_z = \left(\nabla \times \left[\frac{1}{\varepsilon}\mathbf{J}(\mathbf{x})\right]\right) \cdot \hat{\mathbf{z}}. \tag{4.6}$$

where

$$\Lambda = \begin{pmatrix} \frac{1}{1+\mathrm{i}\sigma_x(\mathbf{x})/\omega} & & \\ & \frac{1}{1+\mathrm{i}\sigma_y(\mathbf{x})/\omega} & \\ & & \frac{1}{1+\mathrm{i}\sigma_z(\mathbf{x})/\omega} \end{pmatrix}. \tag{4.7}$$

The PML conductivity $\sigma_\ell(\mathbf{x})$, $\ell = x, y, z$ function is used to gradually "turn on" the PML to compensate for discretization errors [107], and we use a quadratic profile $\sigma_\ell(\mathbf{x}) = \sigma_0(x_{\mathrm{PML}}/d_{\mathrm{PML}})^2$ (where $x_{\mathrm{PML}} \in [0, d_{\mathrm{PML}}]$ is the distance inside the PML).

The exact form of the optimization problem and the corresponding adjoint analysis depend on the specific problem, as will be discussed in more detail in the next section.

4.3   Numerical examples

In this section, we present three example problems in 2D illustrating how our trace-optimization procedure works in practice for typical problems involving ensembles of spatially incoherent emitters. We start in Sec. 4.3.1 with a general case where we are maximizing the total emitted power from many emitters distributed throughout a "fluorescent" dielectric material. Next, in Sec. 4.3.2, we study the enhanced emission from a corrugated surface, analogous to a light-emitting diode [32, 52], showing how

the trace formulation can be applied to a periodic structure with aperiodic emitters. Both of these examples are based on the general algorithm from Sec. 2.7, which can handle emission into a continuum of possible angles. Finally, in Sec. 4.3.3 we apply the more specialized algorithm from Sec. 2.5 to optimizing emission from a fluorescent material into a single-mode waveguide. Since Maxwell's equations are scale-invariant [55], the same optimal designs will be obtained for any wavelength λ if the geometry (thickness and period) is scaled with λ (for the same dielectric constants).

4.3.1 Fluorescent particle

In this example, illustrated in Fig. 4-2a, we optimize the shape/topology of a 2D fluorescent dielectric ($\varepsilon = 12$) particle constrained to have a given area lying within a circular design domain of radius r, maximizing the total power P radiated outwards in *any* direction at a wavelength λ. The emitters are distributed uniformly within the dielectric material.

The governing equation is exactly Eq. (4.6) with $\mu = 1$, whose weak form is [54]:

$$
\begin{aligned}
a(u,v) &= b(v), \\
a(u,v) &= \int_\Omega (\nabla \Lambda v \cdot \frac{1}{\varepsilon} \Lambda \nabla u - k_0^2 vu)\mathrm{d}\Omega\,, \\
b(v) &= \int_\Omega vf\mathrm{d}\Omega\,,
\end{aligned}
\tag{4.8}
$$

where $k_0 = \omega/c$ is the free-space wave number, $f = (\nabla \times \mathbf{J}) \cdot \hat{\mathbf{z}}$ is the source term, and $\nabla \Lambda$ denotes the linear operator $\nabla \Lambda u = \nabla(\Lambda u)$. The matrix A and the source vector $\mathbf{b}$ for the discretized Maxwell equation Eq. (2.2) are obtained by replacing u and v with the finite-element basis functions $\hat{u}_n$ and $\hat{v}_n$, using first-order Lagrange elements on a triangular mesh [54]. The mesh was generated with Gmsh [36], corresponding to a spatial resolution of roughly $\lambda/40$ in the air and $\lambda/80$ in the design region.

Notice that in Eq. (2.22), only U (via A) and B (describing emission only in the dielectric) depend on the design parameters ρ. We have now the optimization

problem as:

$$
\begin{aligned}
g(\rho, W) &= \max_{\rho, W} \operatorname{tr}\left[U(\rho)^\dagger O U(\rho)(W^\dagger B(\rho)W)^{-1}\right], \\
U(\rho) &= A(\rho)^{-1}B(\rho)W, \\
0 &\leq \rho \leq 1, \\
\int \rho \, \mathrm{d}\Omega_d &< \int R_f \mathrm{d}\Omega_d,
\end{aligned}
\tag{4.9}
$$

where we have defined the function $g(\rho, W) = \langle P \rangle$ for the parameter and W dependence, R_f is the area filling-ratio.

Applying adjoint-method analysis [98, 142], we obtain the partial derivatives:

$$
\begin{aligned}
\frac{\partial g}{\partial \rho} = -\operatorname{tr}\left[U^\dagger O U (W^\dagger B W)^{-1}(W^\dagger \frac{\partial B}{\partial p}W)(W^\dagger B W)^{-1}\right] \\
- 2\operatorname{Re}\left\{\operatorname{tr}\left[Z^\dagger\left(\frac{\partial A}{\partial \rho}U - \frac{\partial B}{\partial \rho}W\right)\right]\right\},
\end{aligned}
\tag{4.10}
$$

where Z is the result of an adjoint solve:

$$
A^\dagger Z = O U (W^\dagger B W)^{-1}.
\tag{4.11}
$$

The partial derivative with respect to W is simply obtained via matrix [115] CR calculus [70]:

$$
\frac{\partial g}{\partial W} = \left[I - BW(W^\dagger B W)^{-1}W^\dagger\right](A^{-1}B)^\dagger O U (W^\dagger B W)^{-1}.
\tag{4.12}
$$

We validated the derivatives from the adjoint method against finite differences at random points, and found that the relative error was only about 10^{-6} or less, which is not a problem for the CCSA algorithm when converging the optimum to only a few decimal places.

The analysis workflow for this example is shown in Fig. 4-1. This CCSA update is
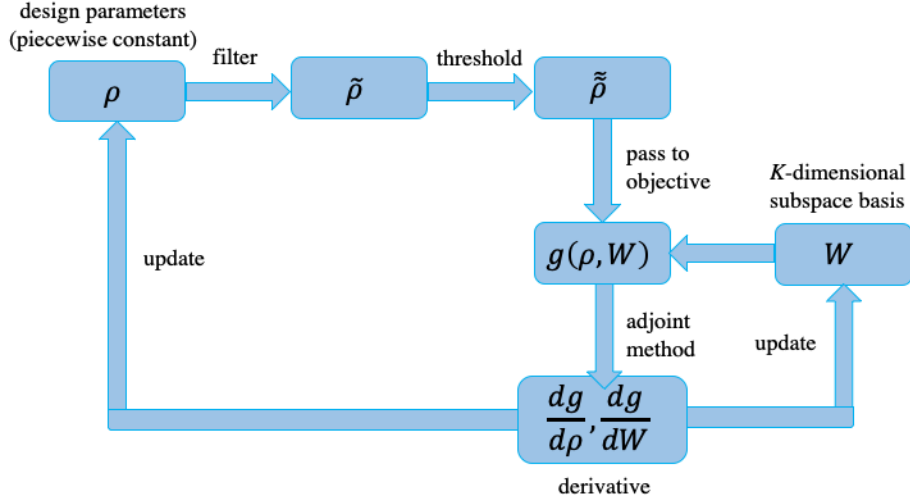
Figure 4-1: Flowchart of the optimization steps for the fluorescent particle and periodic emitting surface examples.

implemented with NLopt in Julia [59] for an increasing series of $\beta = 5, 10, 20, 40, 80$. And for each β, the loop is terminated either a relative difference of 10^{-8} is achieved or the maximum iteration reaches 200. The design parameter ρ is bounded from 0 to 1.

Because this is a non-convex optimization problem, topology optimization can converge to different local optima from different initial geometries [98]. Figure 4-2b shows multiple local-optima geometries for a design radius $r = 0.5\lambda$ with filling-ratio $R_f = 0.5$ and bandwidth quality factor $Q = 1000$ (artificial loss, from Sec. 4.1), obtained from different initial geometries (disks of different radii and/or ε). The numbers above the geometries denote the corresponding emitted (average) power P in arbitrary units. In this particular case, after examining a large number of local optima (not shown), we found that the best local optimum is simply a circular disk with a particular radius. The existence of many local optima with performance varying by factors of 2–5 is not unusual in wave problems [12, 28, 159], and while various heuristic strategies have been proposed to avoid poor local minima [1, 12, 102, 133] beyond simply probing multiple random starting points, the only way to obtain rigorous guarantees is to derive theoretical upper bounds [95, 159] as discussed further in Sec. 4.4 (purely numerical global search can generally provide practical guarantees
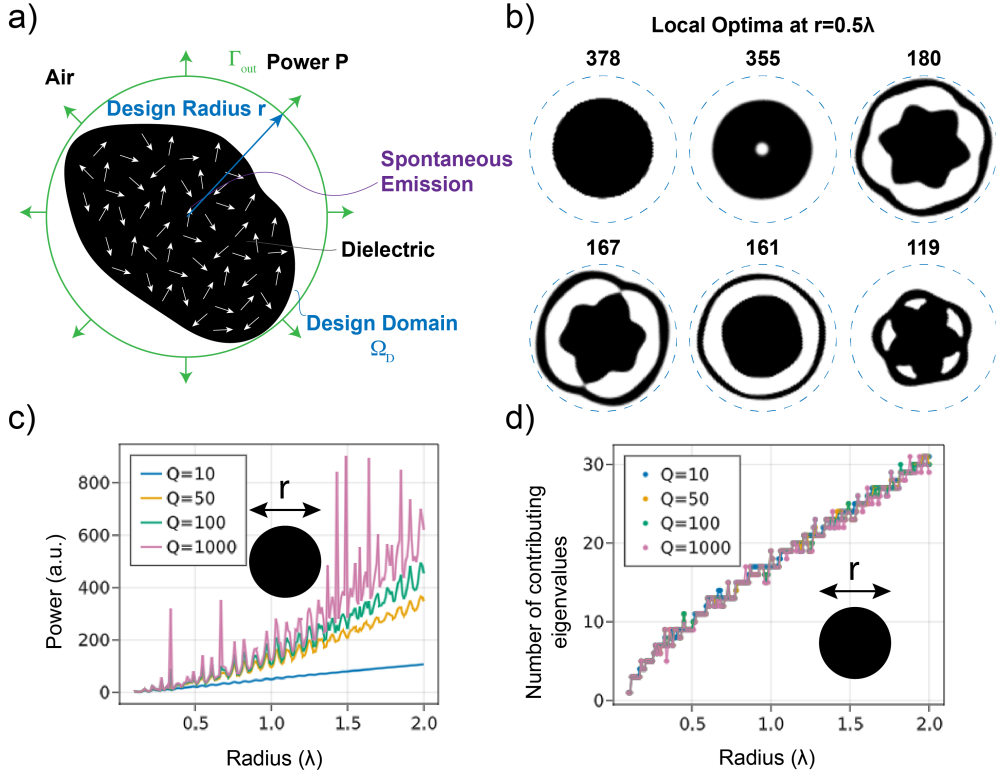
Figure 4-2: (a) A 2D fluorescent particle (of dielectric $\varepsilon = 12$) with a circular design domain of radius r. The emitters are distributed uniformly within the dielectric material. The total power P radiated outwards in *any* direction (integral of Poynting flux over Γ_{out}) at a wavelength λ is optimized. (b) Typical local optima found for design radius $r = 0.5\lambda$ with filling-ratio $R_f = 0.5$ and bandwidth quality factor $Q = 1000$. The numbers above denote the optimized emitting (average) power in arbitrary units. (c) Emitted power of a disk as a function of the disk radius r for different bandwidth quality factors Q. (d) The number of eigenvalues that contribute 99% of the trace as a function of the disk radius r for different bandwidth quality factors Q.

only for very low-dimensional Maxwell optimization [5]).

Whether the best optimum is a disk changes with the design-domain radius, and appears to depend on whether there is a nearby radius with a high-Q resonance at the design λ. (In fact, for this particular case the locally optimal disk has an area slightly less than our upper bound, meaning that the area constraint is not active. In consequence, this particular disk remains a *local* optimum even if the design domain is enlarged, and apparently remains a global optimum until the design domain is sufficiently enlarged to admit a stronger resonance. Although the area constraint is

not active at this particular local optimum, it is active at intermediate points during the optimization process, and there are many *other* local optima that would also be found if the area constraint were not present. Physically that emitted power can increase simply by adding more fluorescent material; correspondingly, without an area constraint we often find a local optimum in which the design region is almost entirely filled with dielectric.) In Fig. 4-2c, we show how the average power radiated by a circular disk varies with radius r/λ, and clearly exhibits a series of sharp peaks correspond to radii which support high-Q resonances at λ: the familiar whispering-gallery resonant modes [156].

The key assumption of our algorithm in Sec. 2.7 was that only a small number of eigenvalues would contribute to the trace, and this assumption clearly holds here. In Fig. 4-2d, we plot the number of eigenvalues that contribute 99% of the trace as a function of the disk radius. We can see that only a small number of eigenvalues is required to obtain a good estimate of the trace; we find similar results for other shapes. Naively, one might expect that the number of contributing eigenvalues would scale with the area (or volume in 3d), corresponding to the number of resonances per unit bandwidth from the density of states (DOS) [160]. However, we find that the scaling is nearly linear with the disk radius; the reason the simple DOS argument fails is that it doesn't take into account the variable loss (radiation) rates of the modes, which causes most of the resonances to contribute weakly even if the *real* part of their frequency is close to the emission frequency. In fact, we have found similar linear scaling of the number of contributing eigenvalues for many other shapes, including other locally optimized shapes, and it appears to be an interesting open theoretical question to prove (or disprove) asymptotic linear scaling.

4.3.2 Periodic emitting surface

In this example, we enhance the emission from a thin "emitting layer" by optimizing a periodically patterned surface situated on top of the layer.—this is inspired by a light-emitting diode (LED) with a patterned surface above an active emitting layer, where it is well known that a periodic pattern can enhance emission via guided-

mode resonances [32, 52, 103]. As illustrated in Fig. 4-3a, the design domain consists of dielectric material ($\varepsilon = 12$) in air with a period L and thickness $H_d = 0.5\lambda$, the spontaneous-emission current sources are uniformly distributed on an horizontal line ("active layer") inside a lower-index substrate ($\varepsilon = 2.25$) a distance $H_s = 0.1\lambda$ below the design domain. The objective, here, is the total power emitted upwards, integrated over *all angles* (i.e., the total Poynting flux) using the methods of Sec. 2.7. (Emission purely into the normal direction could be optimized much more efficiently using the methods of Sec. 2.5.)

We simulate a single unit cell with Bloch-periodic boundary conditions in x. Since Gridap [6] only supports periodic boundary conditions in its current version, we make a change of variables $H_z \to H_z e^{ikx}$ so that H_z is the periodic "Bloch envelope" function [55]. In comparison to Eq. (4.5), this corresponds to the transformation $\nabla \to \nabla + ik\hat{\mathbf{x}}$ [55]:

$$\left[-(\nabla + ik\hat{\mathbf{x}}) \cdot \frac{1}{\varepsilon}(\nabla + ik\hat{\mathbf{x}}) - k_0^2 \right] H_z = f \,, \tag{4.13}$$

with periodic boundaries in x, whose weak form (including PML in y) can then be obtained via integration by parts:

$$
\begin{aligned}
a(u, v) &= b(v), \\
a(u, v) &= \int_\Omega \left[(\nabla\Lambda - ik\hat{\mathbf{x}}) \, v \cdot \frac{1}{\varepsilon} \cdot (\Lambda\nabla + ik\hat{\mathbf{x}}) \, u - k_0^2 v u \right] d\Omega, \\
b(v) &= \int_\Omega v f \, d\Omega \,.
\end{aligned}
\tag{4.14}
$$

where Λ is the diagonal PML "stretching" matrix Eq. (4.7).

The objective (average power) is then constructed by a Brillouin-zone integration over the Bloch wavevector k [103]:

$$g(\rho) = \frac{L}{2\pi} \int_{-\pi/L}^{\pi/L} \text{tr} \left[\left(A_k^{-1} D \right)^\dagger O \left(A_k^{-1} D \right) \right] dk \,, \tag{4.15}$$

61

where L is the period of the unit cell and A_k is assembled using Eq. (4.14). Since this integrand is a periodic function of k, the integral can be approximated by a simple trapezoidal sum over equally spaced points k with exponential accuracy [144]; we used 100 k points in order to resolve sharp resonances.

Commuting the integral and the trace in Eq. (4.15), similarly to Sec. 2.8 (noting that $\int \mathrm{tr} = \mathrm{tr} \int$), we obtain

$$
\begin{aligned}
g(\rho, W) &= \max_{\rho, W} \frac{L}{2\pi} \int_{-\pi/L}^{\pi/L} \mathrm{tr} \left[U_k(\rho)^\dagger O U_k(\rho) (W^\dagger B(\rho) W)^{-1} \right] \mathrm{d}k \,, \, \\
U_k(\rho) &= A_k(\rho)^{-1} B(\rho) W \,, \\
0 &\leq \rho \leq 1 \,.
\end{aligned}
\tag{4.16}
$$

The adjoint analysis for Eq. (4.16) is almost the same as in Sec. 4.3.1, except for the additional integration over k. Also, it shares the same analysis workflow as in Sec. 4.3.1.

Even though the dielectric structure is periodic (the design domain is a single unit cell of ε), the emitters are *not* periodic—they are independent random currents at every point in the active layer. Computationally, however, we can still reduce the simulation of *non-periodic* sources in a periodic medium to a set of small *unit-cell simulations*, using the "array-scanning method" [17]. An arbitrary aperiodic source current can be Fourier-decomposed into a superposition of Bloch-periodic sources $(\mathbf{J}_k(x + L) = e^{ikL}\mathbf{J}_k(x))$, each of which can be simulated with a single unit cell and Bloch-periodic boundary conditions in x. The total power is then simply obtained from an integral $(\int_{-\pi/L}^{\pi/L} \mathrm{d}k))$ over the Bloch wavevector k in the Brillouin zone. For incoherent aperiodic random sources, *each* of these Bloch-periodic unit-cell calculations is an operator trace (over random currents in the unit cell only) computed by the methods of Sec. 2.4. (Unit-cell calculations for different k values are completely independent and can be performed in parallel. Moreover, the array-scanning method can be viewed as a special case of a reduction using symmetry: for any symmetry group, sources can be decomposed into a superposition of "partner functions" of the
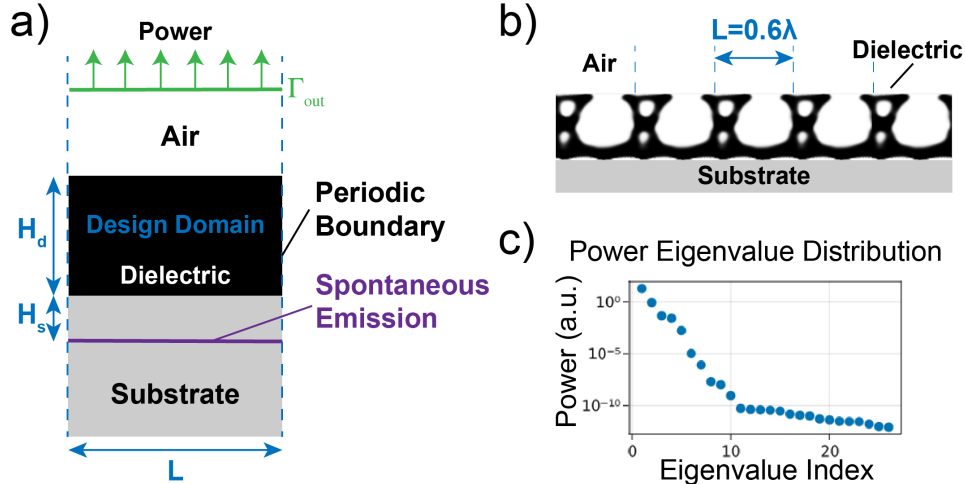
Figure 4-3: (a) Unit cell of a 2D periodic emitting surface with period L. The design domain consists of dielectric material ($\varepsilon = 12$) in air with thickness $H_d = 0.5\lambda$, the spontaneous-emission current sources are uniformly distributed on an horizontal line (purple line) inside a lower-index substrate ($\varepsilon = 2.25$) a distance $H_s = 0.1\lambda$ below the design domain. The objective is the total power emitted upwards, integrated over Γ_{out} (b) Optimized geometry with period $L = 0.6\lambda$. (c) The eigenvalue distribution of the average power for the optimized geometry.

irreducible representations of the symmetry group [50], thus reducing the simulation domain even for asymmetrical random sources.)

The optimized structures for the design parameter $H_d = 0.5\lambda$, $H_s = 0.1\lambda$ is shown in Fig. 4-3b. Note that we have also optimized over the period L (here, simply by repeating the optimization for different values of L) to find an optimized period $L = 0.6\lambda$. The eigenvalue distribution of the average power is given in Fig. 4-3c: Again, we observe that only the first few eigenvalues contribute significantly to the trace, as conjectured in Sec. 2.7.

4.3.3 Emission into a waveguide

This example considers a fluorescent dielectric ($\varepsilon = 12$) medium in air, similar to Sec. 4.3.1, but in this case we are maximizing the power coupled into a single-mode dielectric waveguide ($\varepsilon = 12$, width $\lambda/2\sqrt{12}$) rather than into radiation (Fig. 4-4a). Since the output is a single channel (O is rank 1), this allows us to apply the method of Sec. 2.5 to perform only a single "reciprocal" Maxwell solve per optimization step.
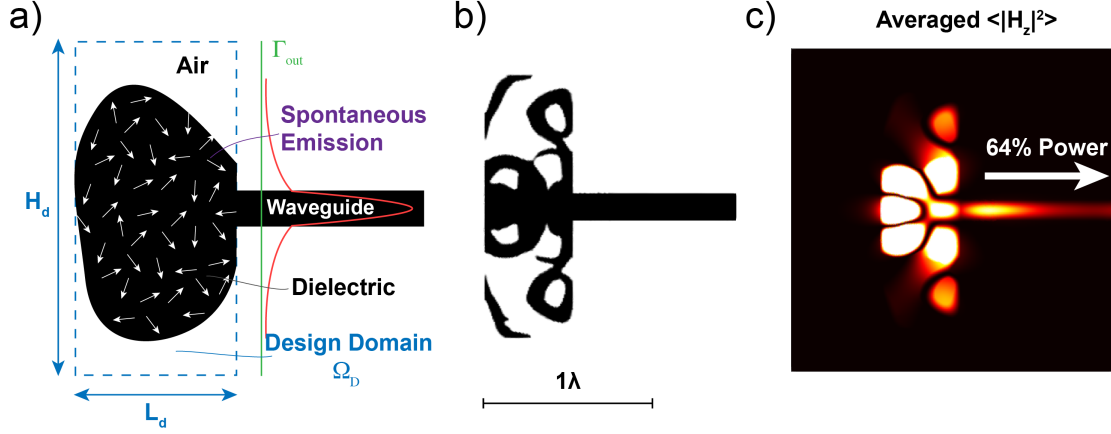
Figure 4-4: (a) A 2D fluorescent dielectric ($\varepsilon = 12$) medium coupling to a waveguide ($\varepsilon = 12$, width $\lambda/2\sqrt{12}$). The design domain is of height $H_d = 1.5\lambda$ and width $L_d = 0.5\lambda$, the power coupled into a single-mode dielectric waveguide (mode overlap integral at Γ_{out}) is optimized. (b) Optimized shape with a filling-ratio $R_f = 0.5$. (c) Averaged field intensity $\langle |H_z|^2 \rangle$ distribution. About 64% percent of the power is coupled into the waveguide mode.

Since the waveguide breaks the rotational symmetry of the problem, the optimum structure is now very different from a circular disk, and must somehow redirect light emitted *anywhere* in the fluorescent material into the waveguide. This task is made more difficult by the fact that we employ a design domain whose size is only $1.5\lambda \times 0.5\lambda$, so the optimization cannot simply surround the emitters with a multi-layer Bragg mirror to confine the radiation (as occurs when optimizing LDOS in a large design domain [81, 150]).

The governing equation and the weak form are identical to Sec. 4.3.1. The main difference is our objective function is now the power in a waveguide mode, computed via an overlap integral using mode orthogonality [136], rather than a total Poynting flux. Here, we briefly review how this overlap integral is implemented in the finite-element method.

For a propagating waveguide mode with electric and magnetic fields $\mathbf{e}_i$ and $\mathbf{h}_i$, the modal-expansion coefficient α_i of that mode for a total magnetic field $\mathbf{H}$ is given by the overlap integral [136]

$$\alpha_i^* = \frac{\int \mathbf{e}_i \times \mathbf{H}^* \cdot d\mathbf{S}}{\int \mathbf{e}_i \times \mathbf{h}_i^* \cdot d\mathbf{S}} = \frac{\int e_{yi} H_z^* dy}{\int e_{yi} h_{zi}^* dy}, \tag{4.17}$$

where we have assumed an x-oriented waveguide in 2D and an in-plane electric-field polarization. The power carried by this mode is then simply $|\alpha_i|^2$. In this example, our objective is the power $|\alpha_0|^2$ in a single mode:

$$\langle P \rangle = |\alpha_0|^2 = \left| \frac{1}{N_0} \int e_{y0} H_z^* \mathrm{d}y \right|^2 , \qquad (4.18)$$

where N_0 is the normalization (which can be omitted for optimization) from Eq. (4.17). If H_z is expressed as a linear combination $\sum_n u_n \hat{u}_n$ of finite-element basis functions $\hat{u}_n$, Eq. (4.18) becomes $\|\mathbf{o}^\dagger \mathbf{u}\|^2$ as in Eq. (2.15), where $\mathbf{o}$ has components o_n given by the linear functional

$$o_n = o(\hat{u}_n) = \frac{1}{N_0} \int e_{y0} \hat{u}_n \mathrm{d}y . \qquad (4.19)$$

Computationally, the assembly of $\mathbf{o}$ in finite-element software is equivalent to constructing a right-hand-side (source) vector $\mathbf{b}$.

The optimization becomes:

$$\begin{aligned}
g(\rho) &= \max_\rho \left[\mathbf{u}(\rho)^\dagger B(\rho) \mathbf{u}(\rho) \right] , , \\
\mathbf{u}(\rho) &= A(\rho)^{-\dagger} \mathbf{o} , \\
0 &\leq \rho \leq 1 . \qquad (4.20)
\end{aligned}$$

By the adjoint method, for any K, we obtain the derivatives:

$$\frac{\mathrm{d}g}{\mathrm{d}p} = \sum_{i=1}^K \left\{ \mathbf{u}_i^\dagger \frac{\mathrm{d}B}{\mathrm{d}p} \mathbf{u}_i - 2\,\mathrm{Re}\left[\mathbf{w}_i^\dagger \left(\frac{\mathrm{d}A^\dagger}{\mathrm{d}p} \mathbf{u}_i \right) \right] \right\} , \qquad (4.21)$$

where $\mathbf{w}_i$ solves $A\mathbf{w}_i = B\mathbf{u}_i$ and $\mathbf{u}_i$ solves the reciprocal problem $A^\dagger \mathbf{u}_i = \mathbf{o}_i$ from Eq. (2.21). This derivative is also compared with the finite difference method and a difference of about 10^{-6} is observed. The analysis work flow is provided in Fig. 4-5.

Figure 4-4b shows the optimized geometry with a design domain of height $H_d = 1.5\lambda$ and width $L_d = 0.5\lambda$. The material is constrained to fill at most half of the design domain (to illustrate that we can independently constrain the design region and the design volume); unlike for the disk optimum in Sec. 4.3.1, this area constraint
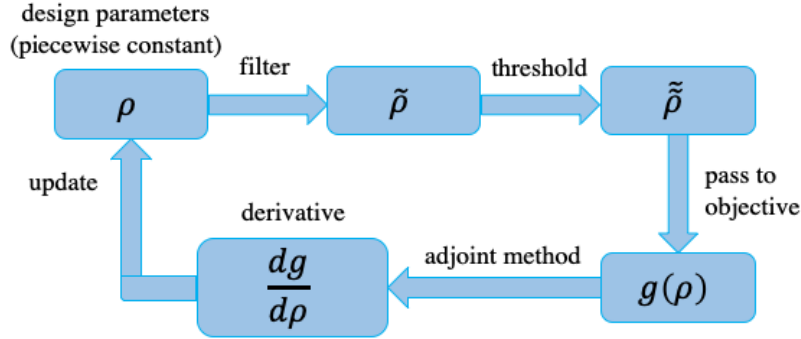
Figure 4-5: Flowchart of the optimization steps for the emission into a waveguide example.

was active at the optimum shown here. The corresponding averaged field intensity $\langle |H_z|^2 \rangle$ is displayed in Fig. 4-4c. We found that 64% of the power is coupled into the desired waveguide mode. In comparison, only 4% of the power is coupled to the waveguide mode for a trivial rectangular design where the the whole design domain is filled with $\varepsilon = 12$ fluorescent material.

4.4 Concluding remarks

In this chapter, we illustrated our trace formulation with several incoherent emission problems in 2D in conjunction with TopOpt. There are still a wide variety of applications that the trace formulation applies. In Chap. 5, we use these techniques to optimize Raman sensing in fluid suspensions of many Raman molecules, in contrast to previous work that only considered a single molecule location [22, 111]—it will help us to answer the interesting open question of the optimal spatial density of "hot spots" where light is concentrated to enhance Raman emission. Another application is enhancing cathodoluminescence or other forms of scintillation detectors, which were previously optimized only for normal emission [130]. In contrast to spon-

taneous emission, where the light is emitted by spatially uncorrelated *point* sources, one can instead consider incoherent beams of light consisting of uncorrelated random *planewave* amplitudes—this corresponds to spatially *correlated* random currents [153], and we are investigating the resulting trace formulation to design metalenses for incoherent focusing. Other applications include the study of radiation loss due to surface roughness, which can be modeled via random sources with a prescribed correlation function related to the manufacturing disorder and may naively require a large number of Maxwell solves [57,66,113]. Nor is our approach limited to Maxwell's equations—it is applicable to any linear system where one wishes to optimize quadratic functions of random source terms.

Algorithmically, we are investigating ways to apply more sophisticated algorithms to the joint structure/trace optimization problem Eq. (2.20). When solving the eigenproblem alone (maximizing over V to obtain extremal eigenvalues), it is well known that one can greatly improve upon straightforward gradient ascent by Krylov algorithms such as Arnoldi [143] or LOBPCG [68], and we would like to incorporate Krylov acceleration into to joint problem as well. Recent techniques to accelerate frequency-domain solves for multiple sparse inputs and outputs [82] may also be applicable to accelerate our trace optimization (since we have multiple sources in a sparse subset of the domain, and objective functions like the power only involve sparse outputs). Similar to the stochastic Lanczos algorithm [145], one could further exploit the fact that we are computing the trace of a function $f(A)$ of the Maxwell operator A in order to relate the trace more efficiently to Krylov subspaces of A. More generally, there are other applications where one is maximizing $\operatorname{tr} f(A(p), p)$ for some f and some parameters p, and it seems similarly beneficial to combine the trace estimation with the parameter optimization in such problems.

Theoretically, it is desirable to complement improved numerical optimizations with new rigorous upper bounds on incoherent emission. Significant progress has already been made on bounding thermal-emission processes [94, 100] as well as to absorption [71,95] (related to emission via reciprocity), and many of these techniques should be adaptable to other forms of random emission.

Chapter 5

Topology optimization of spatially averaged Raman scattering

In this chapter, we present a general framework for inverse design of nanopatterned surfaces for *spatially averaged* Raman enhancement from emitters distributed randomly throughout a material or fluid (Sec. 5.2), building upon the trace formulation proposed in Chap. 2. This leads to radically different designs than optimizing Raman emission at a single known location, as we illustrate using several 2D design problems (Sec. 5.3) addressing effects of hot-spot density, angular selectivity, and nonlinear damage. Looking forward, we believe that this framework sets the stage for future work (Sec. 5.4) in 3D (where field singularities are stronger), TopOpt for dielectric Raman [3] (instead of metal, trading sharper resonances for weaker localization), and related problems in scintillation detectors (where previous work optimized emission but not absorption [130]).

5.1 Introduction

Surface-enhanced Raman scattering (SERS) [34, 75] is a common method to increase the sensitivity of Raman spectroscopy [60] (with enhancements reaching $\sim 10^{10}$ for a single molecule [44, 78, 154, 162]), important for a wide variety of sensing applications [67]. In SERS, active Raman molecules are placed in the vicinity of a textured

surface (e.g. coated with metal nanoparticles [75]) that provides two (multiplicative) resonant enhancements: it concentrates the incoming pump field (at frequency ω_1) at the Raman-molecule location, and it also Purcell-enhances the emission (at a shifted frequency ω_2). In previous work, we derived general upper bounds on the Raman enhancement for arbitrary-shaped structures given the material's susceptibility, the size of the scatterer, and the distance to the Raman molecule [90]. Motivated by optimistic results from these bounds, we used TopOpt [22, 111] to inverse-design novel structures maximizing the Raman enhancement, leading to $\sim 100\times$ improvement over conventional structures [111]. However, this previous work only analyzed the emission of a single Raman molecule placed at a "hot spot" of maximal electric-field intensity [22, 111]. In many practical experiments, however, the Raman molecules are distributed randomly in space, either suspended in a fluid or deposited onto the surface [89, 118]—only a small fraction of the Raman molecules experiences the peak hot-spot enhancement [33, 88, 157]–and it is an open question to determine what structures maximize *average* enhancement over all molecule locations. Some authors have analyzed the effect of one or two geometric parameters on averaged enhancement using a simplified metric discussed below [120, 137], but neither large-scale optimization (e.g. TopOpt) nor a comprehensive theoretical approach have been developed. Also, additional nonlinear effects arise in UV Raman spectroscopy, where extremely high intensities ("too hot") can damage the Raman molecules and suppress emission [157], but the implications of the effect on optimal design have never been analyzed. Spatially distributed Raman emission is challenging to rigorously model, as it naively requires running a large number of simulations for molecules at different locations, which is especially problematic for inverse design where many structures must be simulated over the course of optimization. Building on the trace formulation for optimizing incoherent emission processes proposed in Chap. 2, in this chapter we propose an efficient technique for simulation and inverse design of spatially averaged Raman enhancement and analyze its results for TopOpt applied to several example problems addressing effects of hot-spot density, angular selectivity, and nonlinear damage.

5.2 Model formulation

In this section, we provide a general mathematical framework for optimizing spatially averaged Raman enhancement. We begin with the numerical model for Raman scattering (Sec. 5.2.1) and then show how the trace formulation can be applied to the Raman problem in general (Sec. 5.2.2). Next, we consider the special case where the Raman signals are received in a single direction (Sec. 5.2.3). Finally, we provide some analysis of the singularities in the Raman problem (Sec. 5.2.4).

5.2.1 Numerical model for Raman scattering

The Raman scattering can be modeled as a combination of two electromagnetic processes [75]: first, an incident field (or equivalent current [108] source $\mathbf{J}_1$) produces an electric field $\mathbf{E}_1 e^{-i\omega_1 t}$ at a frequency ω_1. This solves the linear Maxwell equations $M_1\mathbf{E}_1 = i\omega_1\mathbf{J}_1$ where M_1 is the Maxwell (vector Helmholtz) operator $M_1 = \nabla \times \mu_1^{-1}\nabla \times -\frac{\omega_1^2}{c^2}\varepsilon_1$ with $\varepsilon_1(\mathbf{x})$ and $\mu_1(\mathbf{x})$ being the relative electric permittivity and magnetic permeability at frequency ω_1; second, a molecule at position $\mathbf{x}_0$ with a Raman polarizability tensor α produces a dipole current density $\mathbf{J}_2 = \alpha\mathbf{E}_1(\mathbf{x}_0)\delta(\mathbf{x} - \mathbf{x}_0)e^{-i\omega_2 t}$ at a frequency ω_2, which produces an emission field $\mathbf{E}_2$ satisfying $M_2\mathbf{E}_2 = i\omega_2\mathbf{J}_2$ where M_2 is the Maxwell operator at the frequency ω_2. The difference $|\omega_2 - \omega_1|$ is the Raman shift, and usually $|\omega_2 - \omega_1| \ll |\omega_1|$.

Numerically, we discretize this problem (e.g. using finite elements) into a sequence of finite-size linear equations:

$$M_1\mathbf{u}_1 = \mathbf{b}_1 \,, \mathbf{b}_2 = A\mathbf{u}_1 \,, M_2\mathbf{u}_2 = \mathbf{b}_2 \,, \tag{5.1}$$

where $\mathbf{u}_1/\mathbf{u}_2$ is a vector representing the discretized incident/emission fields, A is the discretized Raman polarizability tensor, and $\mathbf{b}_1/\mathbf{b}_2$ is a vector representing the discretized source term. In the following, it is algebraically convenient to work with such a discretized (finite-dimensional) form, to avoid cumbersome infinite-dimensional linear algebra, but one could straightforwardly translate to the latter context as well [55].

Typically, we are interested in maximizing the power radiated into one or more directions/channels by $\mathbf{u}_2$ for a given incident source $\mathbf{b}_1$. It can be expressed as quadratic functions of the emission fields $\mathbf{u}_2$ via the Poynting flux. Since the power is always a real-valued quantity, it corresponds in particular to a Hermitian quadratic form as Eq. (2.3).

When the Raman molecules are distributed randomly in some region, one needs to solve for the emission field $\mathbf{u}_2$ for every single Raman molecule (different α and A) and then take the the average:

$$\langle P \rangle_\alpha = \langle \mathbf{u}_2^\dagger O \mathbf{u}_2 \rangle_\alpha = \langle \mathbf{u}_1^\dagger A^\dagger M_2^{-\dagger} O M_2^{-1} A \mathbf{u}_1 \rangle_\alpha \,, \tag{5.2}$$

where $\langle \cdots \rangle_\alpha$ denotes an average over all allowed molecule positions $\mathbf{x}_0$ and orientations of the Raman molecule (possibly weighted by some nonuniform probability distribution); note that the only terms that depend on the Raman molecules are A and $A^\dagger$. Naively, this average could be computed by a multidimensional quadrature (numerical integral) of Raman solves—that is, we solve Eq. (5.1) for many different positions and orientations in order to explicitly average. However, this could be computationally expensive because of the many Maxwell solves that are required, and may be prohibitive in the context of shape optimization where the averaging must be repeated for many geometric shapes. Instead, we employ the trace formulation proposed in Chap. 2 to efficiently compute this average during shape optimization.

5.2.2   Trace formulation for Raman scattering

The key idea is to rewrite our scalar objective Eq. (5.2) as a "1×1" trace, and then employ the cyclic-shift trace property [76] to group the $A\mathbf{u}_1$ terms together:

$$\langle P \rangle_\alpha = \operatorname{tr}\left[\langle \mathbf{u}_1^\dagger A^\dagger M_2^{-\dagger} O M_2^{-1} A \mathbf{u}_1 \rangle_\alpha \right] = \operatorname{tr}\left[M_2^{-\dagger} O M_2^{-1} \underbrace{\langle A \mathbf{u}_1 \mathbf{u}_1^\dagger A^\dagger \rangle_\alpha}_{B} \right]. \tag{5.3}$$

We now derive a simple, tractable expression for the correlation matrix $B = \langle A\mathbf{u}_1\mathbf{u}_1^\dagger A^\dagger\rangle_\alpha$ arising in the Raman trace (noting that $\mathbf{u}_1 = M_1^{-1}\mathbf{b}_1$ is a fixed vector independent of A). Recall that, for a single Raman molecule at position $\mathbf{x}_0$, the term $\mathbf{b}_2 = A\mathbf{u}_1$ represents the source current generated by the Raman polarizability tensor, discretized in a particular numerical scheme for Maxwell's equations. In particular we consider the expansion of the Raman source current in a finite-element basis [54]:

$$(\mathbf{b}_2)_n = \int_\Omega [\alpha\mathbf{E}_1\delta(\mathbf{x} - \mathbf{x}_0)] \cdot \hat{\mathbf{u}}_n(\mathbf{x})\mathrm{d}\Omega, \tag{5.4}$$

where $\hat{\mathbf{u}}_n(\mathbf{x})$ is the real vector-valued finite element basis function [54] (Nedelec elements in 3D, or $\hat{v}_n\hat{\mathbf{z}}$ with scalar Lagrange elements $\hat{v}_n$ in 2D for z-polarized fields), and Ω is the computational domain. We can then write the components of the correlation matrix B as

$$B_{mn} = \langle(\mathbf{b}_2)_m(\mathbf{b}_2)_n\rangle_\alpha = \iint \hat{\mathbf{u}}_m(\mathbf{x})^T C_\alpha(\mathbf{x}, \mathbf{x}')\hat{\mathbf{u}}_n(\mathbf{x}')\mathrm{d}\Omega\mathrm{d}\Omega', \tag{5.5}$$

where $C_\alpha(\mathbf{x}, \mathbf{x}') = \langle\alpha(\mathbf{x})\mathbf{E}_1(\mathbf{x})\mathbf{E}_1^\dagger(\mathbf{x}')\alpha^\dagger(\mathbf{x}')\rangle_\alpha = \langle\alpha(\mathbf{x})\mathbf{E}_1(\mathbf{x})\mathbf{E}_1^\dagger(\mathbf{x})\alpha^\dagger(\mathbf{x})\rangle_\alpha\,\delta(\mathbf{x} - \mathbf{x}')$ because the emission process is incoherent (different points in space emit with uncorrelated phases).

The simplest case is that of isotropic (scalar) Raman polarizability α, in which case $\langle\alpha(\mathbf{x})\mathbf{E}_1(\mathbf{x})\mathbf{E}_1^\dagger(\mathbf{x})\alpha^\dagger(\mathbf{x})\rangle_\alpha = \langle|\alpha(\mathbf{x})|^2\rangle_\alpha\mathbf{E}_1(\mathbf{x})\mathbf{E}_1^\dagger(\mathbf{x})$. Defining $|\alpha_0(\mathbf{x})|^2 = \langle|\alpha(\mathbf{x})|^2\rangle_\alpha$ as the mean-square polarizability at each point (i.e. the Raman polarizability multiplied by the probability of the molecule being at that point), we obtain $C_\alpha(\mathbf{x}, \mathbf{x}') = |\alpha_0(\mathbf{x})|^2\mathbf{E}_1(\mathbf{x})\mathbf{E}_1^\dagger(\mathbf{x}')\delta(\mathbf{x} - \mathbf{x}')$ and consequently:

$$B_{mn} = \int |\alpha_0(\mathbf{x})|^2\hat{\mathbf{u}}_m^T\mathbf{E}_1\mathbf{E}_1^\dagger\hat{\mathbf{u}}_n\mathrm{d}\Omega, \tag{5.6}$$

For more the general case of an anisotropic Raman polarizability tensor α, the expression $\langle\alpha(\mathbf{x})\mathbf{E}_1(\mathbf{x})\mathbf{E}_1^\dagger(\mathbf{x})\alpha^\dagger(\mathbf{x})\rangle_\alpha$ must be averaged over all possible *orientations* of the Raman molecule. If we assume that all orientations are equally likely, the

average can be computed explicitly [101], resulting in:

$$B_{mn} = \int \hat{\mathbf{u}}_m^T \left[(\alpha_\parallel^2 - \alpha_\perp^2) \mathbf{E}_1 \mathbf{E}_1^\dagger + \alpha_\perp^2 \mathbf{E}_1^\dagger \mathbf{E}_1 I \right] \hat{\mathbf{u}}_n \mathrm{d}\Omega \,, \tag{5.7}$$

where I is the 3×3 identity matrix and

$$\alpha_\parallel^2 = \frac{|\operatorname{tr}\alpha|^2 + \operatorname{tr}[\alpha^\dagger \alpha + \alpha^* \alpha]}{15} \,, \qquad \alpha_\perp^2 = \frac{\operatorname{tr}[\alpha^\dagger \alpha]/3 - \alpha_\parallel^2}{2} \tag{5.8}$$

(where α^* denotes the complex conjugate).

Once the correlation matrix B is determined, we can then apply different techniques developed in Sec. 2.4 to combine the trace estimation problem with the shape optimization for different scenarios depending on the number of input and output channels. In this work, we focus on the case where the emitted Raman signals are received in a single direction/channel, which means the objective matrix O now becomes rank 1, and the trace formulation also simplifies to two Maxwell solves (one forward and one reciprocal) as discussed below.

5.2.3 Single-channel simplification

For spatially incoherent Raman emission into a single direction/channel, the average power of all the emitters can be computed with a single "reciprocal" solve. The power emitted into a single direction can be expressed as a *mode overlap integral* of the emitted field $\mathbf{u}_2$ and a planewave mode $\mathbf{o}$, which is algebraically of the form $\|\mathbf{o}^\dagger \mathbf{u}_2\|^2$ [136]. In terms of the quadratic form Eq. (2.3), the objective matrix O is now simply a rank-1 matrix $O = \mathbf{o}\mathbf{o}^\dagger$, and the objective trace Eq. (5.3) reduces to

$$\langle P \rangle_\alpha = \operatorname{tr}\left[M_2^{-\dagger} \mathbf{o}\mathbf{o}^\dagger M_2^{-1} B \right] = \mathbf{u}_2'^\dagger B \mathbf{u}_2' \,., \tag{5.9}$$

where $\mathbf{u}_2' = M_2^{-\dagger}\mathbf{o}$ corresponds to solving a (conjugate-) *transposed* Maxwell problem with a "source" $\mathbf{o}$ at the *output* location. Note that matrix B is constructed from the pump field $\mathbf{u}_1$, so Eq. (5.9) requires only two Maxwell solves (pump field $\mathbf{u}_1 = M_1^{-1}\mathbf{b}_1$

and reciprocal field $\mathbf{u}_2' = M_2^{-\dagger}\mathbf{o}$) to obtain the averaged power.

The formulation can be further simplified when the Raman molecule is isotropic. Inserting Eq. (5.6) into Eq. (5.9), we obtain

$$\langle P \rangle_\alpha = \int |\alpha_0(\mathbf{x})|^2 |\mathbf{E}_1(\mathbf{x})|^2 |\mathbf{E}_2'(\mathbf{x})|^2 \, d\Omega \, , \tag{5.10}$$

where $\alpha_0(\mathbf{x})$ indicates Raman molecules distribution, $\mathbf{E}_1(\mathbf{x})$ is the pump field (constructed from $\mathbf{u}_1$) and $\mathbf{E}_2'(\mathbf{x})$ is the reciprocal field (constructed from $\mathbf{u}_2'$). Therefore the averaged power is just an overlap integral of the Raman molecule distribution, the pump field intensity, and the reciprocal field intensity.

From Eq. (5.10), we can also see that the equation further simplifies if (i) the pump and emission directions are the same and (ii) we make an approximation of a negligible Raman shift ($\omega_1 \approx \omega_2$), in which case one can take $\mathbf{E}_2' \approx \mathbf{E}_1$. For isotropic Raman polarizability whose mean $|\alpha_0|^2$ is constant in some volume V and zero elsewhere, this leads to the $\int_V |\mathbf{E}_1|^4$ figure of merit used in several previous works [120, 137] for Raman power (which is often presented heuristically, but has also been justified using reciprocity [79]).

5.2.4 Corner singularities and hot spots

It is common knowledge that SERS tends to favor geometries with "hot spots" where high field intensities arise from geometric singularities (such as sharp tips/cusps, bowtie antennas, or touching spheres), especially for single-molecule SERS where the enhancement theoretically diverges in the limit of arbitrarily sharp tips (in the continuum macroscopic Maxwell equations with local materials). However, it is less clear whether the average Raman enhancement of many volume-distributed emitters still favors such hot spots, since the effect of a field singularity might be averaged out. In this section, we analyze the effect of corner singularities on average Raman in 2D and 3D for the single-channel isotropic-Raman case of Eq. (5.10).

Field singularities at sharp corners are frequency-independent and can be analyze purely using electrostatics [4], so the pump $\mathbf{E}_1$ and reciprocal $\mathbf{E}_2'$ fields have identical

scaling near a sharp tip. Therefore, without loss of generality, we can analyze the simplified metric $\int |\mathbf{E}_1|^4$ in the neighborhood of a sharp tip.

For a 2D sharp corner in a dielectric or metallic material ε enclosing an angle $\phi < \pi$, the field singularity of the field $\mathbf{E}_1$ is a fractional power law in the distance r from the tip [4]:

$$\mathbf{E}_1 \sim r^{t-1}, \tag{5.11}$$

where the exponent $1/2 < t < 1$ depends on ε and ϕ via a transcendental equation [4], which simplifies to $t = \frac{\pi}{2\pi - \phi}$ for a perfect electric conductor. The contribution of this singularity to $\int |\mathbf{E}_1|^4$ then scales as $\int r^{4(t-1)} r \, dr \sim r^{4t-2}$, but $t > 1/2 \implies 4t - 2 > 0$ so the integral is finite (the singularity is integrable). Hence, in 2D there is no reason for optimization to favor *arbitrarily* sharp tips. (The same is true for "2D edges" in 3D.) In Sec. 5.3.1, we correspondingly show that the topology-optimized geometry does not exhibit sharp features (even without manufacturing constraints to prohibit such features [43]), and performs better than optimized touching spheres or bowtie antennas with a field singularity. (The optimized fields still exhibit "hot spots" with high intensity, but no singularities.)

In 3D, the field singularity at sharp tips (e.g. cones or corners) is stronger than in 2D. For example, the fields at the tip of a 3D cone with angle $\phi < \pi$ also exhibit a singularity $\mathbf{E}_1 \sim r^{t-1}$ but with a stronger power law $0 < t < 1$ (e.g. $t = \frac{1}{2\log(8/\phi)} > 0$ for perfect conductors) [48]. The integral then becomes $\int r^{4(t-1)} r^2 \, dr \sim r^{4t-1}$, which diverges for $t < 1/4$ (sufficiently small ϕ). Therefore, we expect that 3D topology optimization of incoherent Raman emission will favor arbitrarily sharp tips, limited only by the imposition of manufacturing constraints [43].

It is worth contrasting the Raman case, in which the field singularity is *squared* by the conjunction of pump and emission enhancement, with spontaneous emission in cases with non-optical pumping (such as light-emitting diodes [52], scintillation from high-energy particles [130], or thermal emission). In such cases, if the excitation is nearly uniform in the vicinity of a sharp tip, then the emitted power scales as $\int |\mathbf{E}_2'|^2$ from the reciprocal field alone. The contribution of a corner singularity is

then $\sim \int r^{2(t-1)}r\, dr \sim r^{2t}$ in 2D ($t > 1/2$) and $\sim \int r^{2(t-1)}r^2\, dr \sim r^{2t+1}$ in 3D ($t > 0$), both of which vanish as $r \to 0$. In consequence, one does *not* expect arbitrarily sharp corners/tips to be favored when optimizing spatially averaged emission alone (or spatially averaged local density of states, LDOS). Indeed topology optimization of incoherent emission (Sec. 4.3) or scintillation [130] did not exhibit arbitrarily sharp corners, in contrast to the "bowtie antenna" singularities that typically arise when optimizing emission/LDOS from a *single* emitter location [24, 29]. (Similar considerations apply to optimization of photovoltaic cells, since maximizing absorption is equivalent to maximizing spatially averaged emission via Kirchhoff's law.)

5.2.5 Numerical formulation

We follow the TopOpt formulation in Sec. 4.1 but with a different interpretation of the material and a different β series ($\beta = 8, 16, 32$). Here, we describe the material by an electric relative permittivity (dielectric constant) $\varepsilon(\mathbf{x})$ in the Maxwell operator (M_1 or M_2), given by:

$$\varepsilon(\mathbf{x}) = \left[n_{\mathrm{f}} + (n_{\mathrm{metal}} - n_{\mathrm{f}})\tilde{\tilde{\rho}}(\mathbf{x}) \right]^2 , \qquad (5.12)$$

where n_{f} is the refractive index of the background fluid (water, $n_{\mathrm{f}} = 1.33$) and n_{metal} is the complex refractive index of the design metal (silver) throughout this work. Note that we interpolate the electric relative permittivity of the material via the refractive index instead of directly from the electric relative permittivity as Eq. (4.3), in order to avoid artificial singularities that may arise when interpolating between metallic (negative) and dielectric (positive) ε [23].

Similar to Sec. 4.2, we employ the frequency-domain Maxwell equations Eq. (4.5) for the magnetic field $\mathbf{H}$ arising from an in-plane (H_z-polarized) electric current $\mathbf{J}$. The electric field $\mathbf{E}$ is then retrieved from $\mathbf{H}$ by taking the curl of $\mathbf{H}$ [55]:

$$\mathbf{E}(\mathbf{x}) = \frac{\mathrm{i}}{\omega\varepsilon(\mathbf{x})}\nabla \times \mathbf{H}(\mathbf{x}) . \qquad (5.13)$$

77

For the Raman scattering process, similar to Eq. (5.1), one can also derive a series of linear equations but $\mathbf{u}_1$ and $\mathbf{u}_2$ are now the discretized magnetic fields, M_1 and M_2 are now discretized Maxwell operators in the magnetic form, A is also changed which now contains not only the Raman polarizability tensor term but also a magnetic-field to electric-field conversion term and a electric-current to equivalent magnetic-current conversion term. We can rewrite Eq. (5.6) in terms of the scalar magnetic field H:

$$B_{mn} = \int |\frac{\alpha_0(\mathbf{x})}{\varepsilon_1\varepsilon_2}|^2 |\nabla\mathbf{H}_1|^2 (\nabla\hat{\mathbf{u}}_m \cdot \nabla\mathbf{u}_n)\,\mathrm{d}\Omega\,, \tag{5.14}$$

where $\mathbf{H}_1$ is the pump field in magnetic form constructed from $\mathbf{u}_1$.

For simplicity, we denote the objective $\langle P\rangle_\alpha$ by g, Eq. (5.9) still has the form

$$g(\tilde{\rho}) = \mathbf{u}_2'^\dagger(\tilde{\rho})B(\tilde{\rho})\mathbf{u}_2'(\tilde{\rho})\,. \tag{5.15}$$

Via the adjoint method, the derivative of the power with respect to the density ρ can be computed via

$$\frac{\mathrm{d}g}{\mathrm{d}\tilde{\rho}} = \mathbf{u}_2'^\dagger\frac{\mathrm{d}B}{\mathrm{d}\tilde{\rho}}\mathbf{u}_2' - 2\operatorname{Re}\left[\mathbf{w}_2^\dagger\left(\frac{\mathrm{d}M_2^\dagger}{\mathrm{d}\tilde{\rho}}\mathbf{u}_2'\right)\right]\,, \tag{5.16}$$

where $M_2\mathbf{w}_2 = B\mathbf{u}_2'$ is the "adjoint" solve. Note that the correlation matrix B is a function of $\mathbf{u}_1$, so the chain rule for the derivative $\frac{\mathrm{d}B}{\mathrm{d}\tilde{\rho}}$ must be computed with another adjoint solve but in a similar manner.

5.3 Results

In this section, we present various example problems building from a single-channel framework (Sec. 5.2.3), illustrating the key tradeoffs and physical effects. We begin with the simplest case with normal incidence and emission (Sec. 5.3.1), revealing an optimal density of the hot spots. Next, we show that the pump and emission angles can also be considered as design parameters. We provide an example where the pump field is fixed at angle $\theta_1 = 30°$ and we search for the optimal emission angle, which turns out to be roughly normal emission and performs nearly as well as
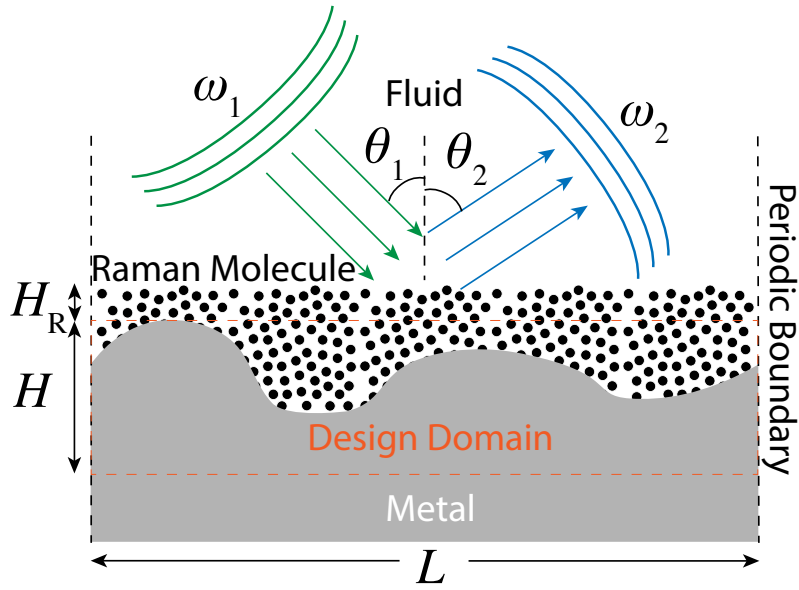
Figure 5-1: Sketch of the 2D Raman scattering design problem. The Raman molecules are distributed uniformly in fluid (water, $n_\mathrm{f} = 1.33$) background near a periodically patterned metal (silver) surface with period L. The incident planewave (H_z-polarized, $\lambda_1 = 532$ nm, green) at angle θ_1 excites the Raman molecules and the Raman-shifted signal ($\lambda_2 = 549$ nm, blue) at angle θ_2 is measured and optimized.

the normal-incidence pumping but with a very different design (Sec. 5.3.2). Then we take into consideration the effect of UV-like nonlinear damage [157] suppressing the emission and optimize with this taken into account, which again leads to very different designs (Sec. 5.3.3). Finally, we briefly discuss the case where Raman emission is only from molecules coating the material *surface* rather than being distributed throughout the volume of a fluid (Sec. 5.3.4).

Figure 5-1 is a sketch of the single-channel Raman scattering design problem in 2D. The Raman molecules are distributed uniformly in a fluid (water) background above a periodically patterned metal (silver) surface with period L. An incident planewave (H_z-polarized, $\lambda_1 = 532$ nm) at angle θ_1 excites the Raman molecules and the Raman-shifted power ($\lambda_2 = 549$ nm) at angle θ_2 is measured and optimized. The design region is a $L \times 200$ nm rectangular domain, in which the material can either be fluid (with Raman molecules) or metal (without Raman molecules). We sweep the optimization over different periods L to find an optimal period (which corresponds
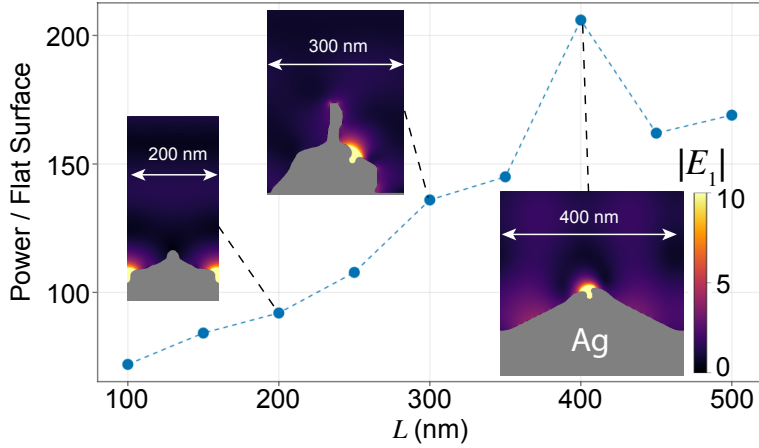
Figure 5-2: Power of the optimized surfaces (compared to that of a flat surface) over different periods for silver at normal pump ($\lambda_1 = 532$ nm, $\theta_1 = 0°$) and emission ($\lambda_2 = 549$ nm, $\theta_2 = 0°$). The insets show the optimized patterns and the pump fields $|E_1|$ in a unit cell for typical periods.

to an optimal density of hot spots). An infinitely thick layer of Raman molecules would emit infinite power in the absence of water absorption and pump depletion, but since we are only interested in optimizing near-field enhancement we limit the Raman molecules to a half-wavelength layer of thickness $(\lambda_1 + \lambda_2)/(2n_\mathrm{f})$ above the design domain.

5.3.1 Normal incidence and emission

In this example we consider the case when the incident pump and measured emission are both normal to the surface ($\theta_1 = \theta_2 = 0°$). We optimize this power P of the emitted Raman signal at different periods L, normalized to a baseline power for a flat metal surface (metal half-space). Since this optimization problem is non-convex, TopOpt can easily converge to different local optima from different initial geometries [98]. In Fig. 5-2, we plot only the largest local optima we found for each period (from 100 nm to 500 nm) for 20 different random starting structures. (We find that the largest local optima have very similar performance, within $\sim 10\%$, giving us some confidence that there are unlikely to be dramatically better local optima yet to be found.)
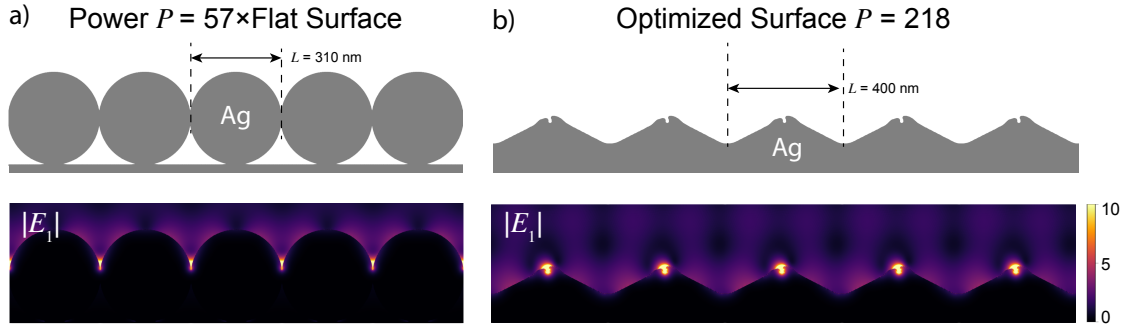
Figure 5-3: Comparison of a) optimized sphere and b) optimized surface for the normal pumping and emission case. The optimized sphere is found to have a diameter of 310 nm and adjacent to each other. The optimized surface is found to have a period of $L = 400$ nm, approximately equal to the pumping and emission wavelength in water. The pumping fields $|E_1|$ are displayed below the pattern (The emission fields $|E_2|$ are similar because of the small Raman shift).

As shown in the inset of Fig. 5-2, the optimized patterns share a small "notch" feature that creates a hot spot (localized resonance) and some of them exhibit spontaneous symmetry breaking (the resulting pattern is asymmetric although the problem is mirror-symmetric)| [30]. This "notch" hot spot is different from the field singularity arising from with sharp corners where the field theoretically diverges. Here, the minimum length-scale of those notches is about 10 nm. The optimal density of those hot spots is found to be at period $L = 400$ nm, which is nearly the wavelength $\lambda_{\text{vacuum}} \sqrt{\frac{1}{\varepsilon_{\text{f}}} + \frac{1}{\varepsilon_{\text{metal}}}} = 375$ nm of surface plasmons at a flat silver–air interface [161]. We also did the same optimization in the UV regime with $\lambda_1 = 400$ nm and $\lambda_2 = 437$ nm (not shown here), and the optimal period $L = 300$ nm was also found to be close to the surface-plasmon wavelength 241 nm. Intuitively, periodic perturbations with this wavelength implement a grating coupler between normal-incident radiation and a surface-plasmon resonance [161], but of course the period changes as the surface is deformed substantially. We also performed an optimization for a doubled period of $L = 800$ nm (not shown), and unsurprisingly found that it converges to two hot spots in each unit cell with similar performance to the single-hotspot $L = 400$ nm design.

To gain a better understanding of how the optimized surface performs, we also
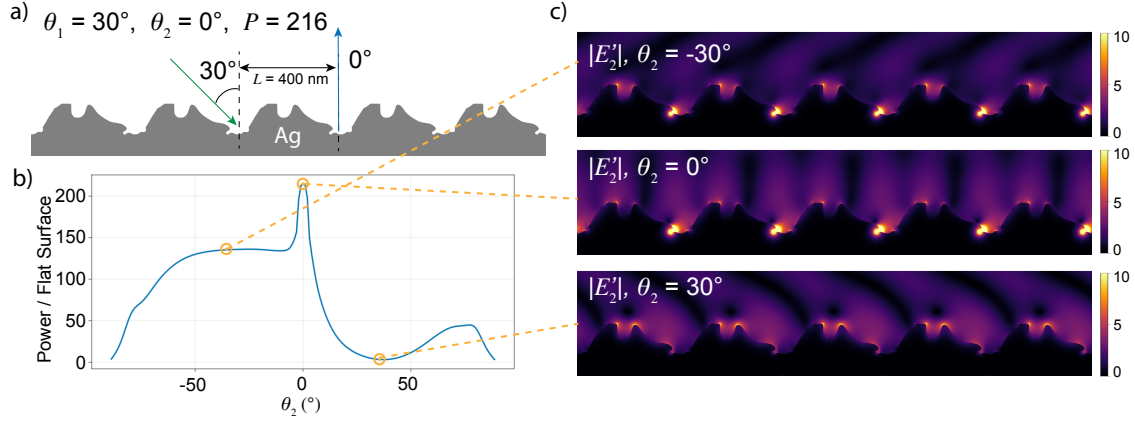
Figure 5-4: Optimized surface with a fixed incident angle at $\theta_1 = 30°$. a) Pattern of the optimized surface, for the optimal emission angle: normal emission ($\theta_2 = 0°$). b) Emitted power (per unit cell) of the optimized surface with $\theta_1 = 30°$ as a function of the emission angle θ_2, normalized by the normal emission of a flat surface. c) Reciprocal fields $|\mathbf{E}_2'|$ at selected emission angles.

compare to a surface coated with optimized spheres (we optimized over both the sphere diameter and period). As shown in Fig. 5-3a, the optimized spheres are of diameter 310 nm and have periods equal to their diameters (i.e. touching). The performance of our TopOpt surface is about 4× better than that of the optimized spheres (A similar optimization was carried out for a bowtie structure, and the optimized bowtie performs slightly worse than the optimized spheres). From the pump field displayed below the pattern in Fig. 5-3, we can also see that the notch hot spot in the optimized surface is more spread out than the singular hot spot produced by tangent spheres—as predicted in Sec. 5.2.4, 2D distributed Raman emission does not favor singularities at points or cusps.

5.3.2 Oblique incidence

In this example (Fig. 5-4), we consider the case where the incident pump field is fixed at an angle $\theta_1 = 30°$ and we search for an optimal emission angle θ_2 that maximizes the power (per unit cell) of the emitted signal (with a fixed period $L = 400$ nm).

In Fig. 5-4a, we show the optimized surface for normal emission, which we find to be optimal by scanning the power over the emission angle θ_2 (Fig. 5-4b). The

structure again has small notches that create hot spots, but with a very different design compared to the normal-incident pump in the previous section. The power near $\theta_2 = -30°$ (inverse to the incident direction) is also large, which is expected since the reciprocal field is similar (for small Raman shift) to the pump field when $\theta_1 = -\theta_2$. On the other hand, the emitted power at $\theta_2 = 30°$ is very low. From Fig. 5-4c, we can see that this is because the reciprocal fields $\mathbf{E}'_2$ excite a localized resonance for both $\theta_2 = -30°$ and $0°$ but not at $\theta_2 = 30°$: the resonant frequencies of a periodic surface depend on angle (corresponding to Bloch wavevector).

5.3.3 Nonlinear damage

In this example, we discuss a nonlinear damage effect that suppresses emission at very high field intensity, similar to phenomena observed experimentally for UV Raman spectroscopy [33, 157]. In particular, we assume that pump fields larger than some threshold magnitude $|\mathbf{E}_{\mathrm{th}}|$ will damage the Raman molecules and inhibit emission.

We model this phenomenon by treating the mean Raman susceptibility $|\alpha_0|^2$ in Eq. (5.10) as *nonlinear*, exponentially decreasing for pump fields larger than the threshold. We replace $|\alpha_0|^2$ with

$$\frac{|\alpha_0(\mathbf{x})|^2}{1 + \exp\left[\gamma(|\mathbf{E}_1|^2 - |E_{\mathrm{th}}|^2)\right]} \, , \tag{5.17}$$

where γ is a coefficient that determines the rapidity of the damage threshold. A sharp cutoff for emission would correspond to $\gamma \to \infty$, but such a step-function behavior would make the problem non-differentiable and impractical to optimize. Instead, we use $\gamma = \gamma_0/|E_{\mathrm{th}}|^2$, where $\gamma_0 \in \{1, 10, 100\}$ is gradually increased during optimization similar to the β parameter of binarization reviewed in Sec. 4.1.

Figure 5-5a shows the optimized surface for a nonlinear damage threshold $|E_{\mathrm{th}}| = 10$ (for an pump planewave with $|E| = 1$). We can see that the pump-field pattern is much more spatially spread out than the highly concentrated hot spots of the previous sections. For comparison, if we use the previous optimized structure (without threshold) from Sec. 5.3.1 with a nonlinear damage, we find that its performance
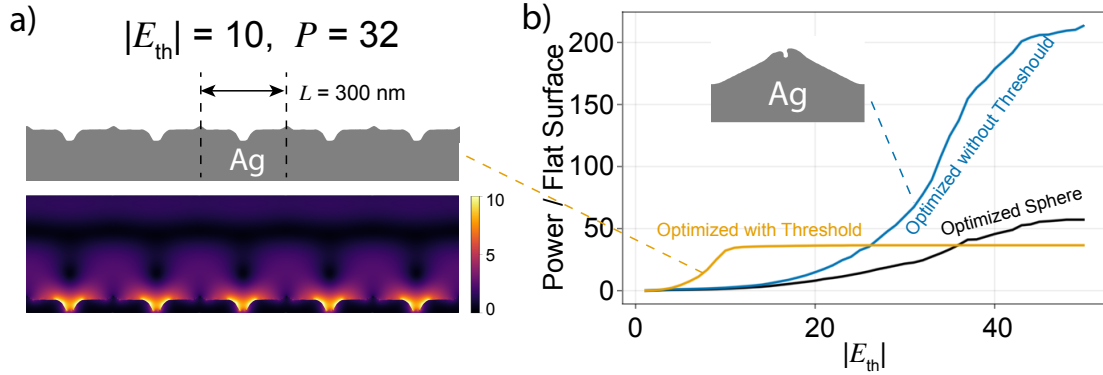
Figure 5-5: a) Pattern of the optimized surface with target threshold $|E_{\mathrm{th}}| = 10$ and the corresponding pump field $|E_1|$. b) Power of optimized surface without threshold (blue), optimized spheres (black) and optimized surface with threshold (orange) as a function of the damage threshold $|E_{\mathrm{th}}|$ (for a pump planewave with $|E| = 1$).

rapidly degrades in the presence of nonlinear damage. This is shown in Fig. 5-5b: the emitted power falls rapidly with the damage threshold for the optimized surface from Sec. 5.3.1 or the optimized spheres (i.e., it is easy for the pump field to damage the Raman molecules), while the design with nonlinear damage taken into account only goes down for thresholds lower than the design target $|E_{\mathrm{th}}| = 10$, and it performs about 20× better than the optimized spheres at $|E_{\mathrm{th}}| = 10$.

5.3.4   Surface-averaged Raman enhancement

In some experiments, there is a monolayer of Raman molecules deposited on the metal surface [89, 157], which increases their exposure to the resonant enhancement. In this case, the average emission should be computed as a integral over the surface instead of the volume as in the examples above. One simple technique to model this is to make the mean Raman polarizability $|\alpha_0|^2$ proportional to $\tilde{\tilde{\rho}}(1 - \tilde{\tilde{\rho}})$ (where $\tilde{\tilde{\rho}}$ is the thresholded and filtered density field from Sec. 4.1), which is ≈ 0 except near the metal surface. This allows us to take the surface geometry into account during sensitivity analysis and optimization. With a finite filter radius $r_f = 2$ nm and a finite $\beta = 32$ smoothing out the structure, we still obtain similar notch-featured patterns (shown in Fig. 5-6) which perform about 3× better than optimized spheres.
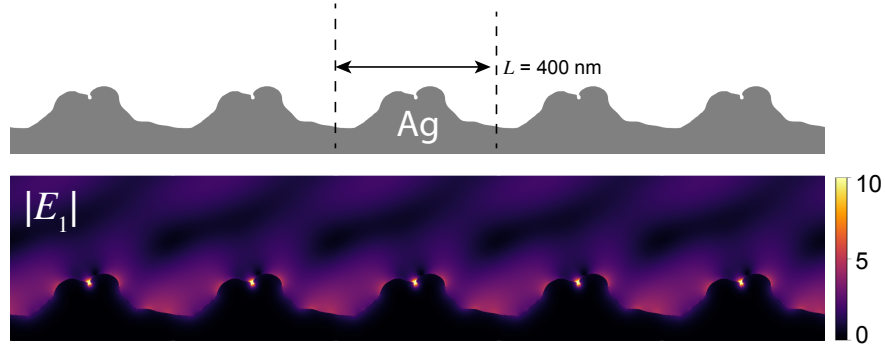
Figure 5-6: Optimized surface for surface-average instead of volume-average at period $L = 400$ nm.

One thing to be noted is that the surface integral could diverge in principle at sharp corners, as can be seen from the $\int |\mathbf{E}_1|^4$ metric (Sec. 5.2.4). Here, the design does not seem to exhibit a sharp corner, probably because we imposed a soft minimum lengthscale by setting a smoothing filter radius of $r_f = 2$ nm. However, ultimately one might need to impose manufacturing constraints and/or nonlinear damage thresholds to prevent inverse design from favoring singular structures for surface emission, even in 2D.

5.4   Concluding remarks

We presented a general framework for optimizing spatially averaged Raman enhancement, which requires only two Maxwell solvers per optimization step for emission in a single direction, and we illustrated this technique for a number of 2D examples. This illustrates the computational technique, and the basic phenomena of optimal hot-spot densities, angular dependence, and the effect of nonlinear damage. The next step is to carry these techniques into 3D, where the same computational principles apply but radically different structures may arise due to the stronger singularities at sharp 3D tips (Sec. 5.2.4). In 3D, these singularities mean that the imposition of manufacturing constraints [43] and/or nonlinear thresholding will play a key role. Because only a small range of geometries have previously been explored for this problem, it is possible that substantial practical improvements may be uncovered by TopOpt

for 3D SERS, especially in less-explored circumstances such as distinct input/output directions, nonlinear damage, or even integrated SERS with waveguide channels [22]. Another important complementary problem is the development of theoretical upper bounds to distributed Raman emission, generalizing earlier work bounding emission at a single location [90], as well as related efforts to bound the "density" of resonant modes (e.g. for solar cells [160]). Computationally, there are a wide variety of nonlinear-optics problems that may potentially be optimized using techniques involving coupled linear Maxwell solves, from scintillation processes [130] to harmonic generation [83], and balancing the tradeoffs between multiple physical processes is precisely where large-scale optimization has the greatest advantages over intuitive human design.

Chapter 6

Concluding remarks

In this thesis, we proposed a trace formulation for inverse design of random emitters, which is much more efficient than direct approaches.

We began with optimizing the LDOS, which is described by a single emitter with random polarization, and showed that theoretical limits are nearly attainable via a metallic-cavity optimization. We then presented a more general case where emitters are distributed randomly in space with known statistics. We showed that only a small number of simulations is required for each optimization step using the trace formulation. In particular, when the output is coupled into a single channel, only one reciprocal simulation is needed to compute the total (average) response, which is closely related to the Kirchhoff's law in thermal radiation. Finally we applied the trace formulation on the spatially averaged SERS enhancement and analyzed its results for TopOpt applied to several example problems addressing effects of hot-spot density, angular selectivity, and nonlinear damage.

There are many future directions that are worth exploring. For the many-output-channel trace formulations, it would be interesting to investigate ways to apply more sophisticated algorithms to accelerate the optimization, a topic we are currently exploring with Dr. Andrew Horning at MIT. Also, recent techniques [82] to accelerate frequency-domain solves for multiple sparse inputs and outputs may be applicable to accelerate our trace optimization (since we have multiple sources in a sparse subset of the domain, and objective functions like the power only involve sparse outputs).

For the spatially averaged SERS enhancement, a 3D design with manufacturing constraints might yield unexpected geometries—it is expected to be qualitatively different than the 2D case due to the change in the nature of the singularity at sharp corners—and we hope to both complete a 3D design and collaborate on an experimental realization. In addition, an upper bound on the spatially averaged SERS enhancement might be derived similar to the single molecule case, which can provide some guidance during the optimization. We also hope to apply an analogous trace-optimization algorithm to other problems such as scintillation, incoherent focusing (recently implemented in Ref. 114), and Casimir effects.

Bibliography

[1] N. Aage and V. Egede Johansen. Topology optimization of microwave waveguide filters. *International Journal for Numerical Methods in Engineering*, 112(3):283–300, 2017.

[2] Mario Agio and Diego Martin Cano. The Purcell factor of nanoresonators. *Nat. Phon.*, 7:674–675, 2013.

[3] Ivano Alessandri and John R Lombardi. Enhanced Raman scattering with dielectrics. *Chemical reviews*, 116(24):14921–14981, 2016.

[4] J Andersen and V Solodukhov. Field behavior near a dielectric wedge. *IEEE Transactions on Antennas and Propagation*, 26(4):598–602, 1978.

[5] Paul Azunre, Joel Jean, Carmel Rotschild, Vladimir Bulovic, Steven G. Johnson, and Marc A. Baldo. Guaranteed global optimization of thin-film optical systems. *New Journal of Physics*, 21:073050, 2019.

[6] Santiago Badia and Francesc Verdugo. Gridap: An extensible finite element toolbox in julia. *J. Open Source Softw.*, 5(52):2520, 2020.

[7] P. K. Banerjee and R. Butterfield. *Boundary element methods in engineering science*. McGraw-Hill Book Co., 1981.

[8] Gang Bao, Yanzhao Cao, Junshan Lin, and Hans Werner Van Wyk. Computational optimal design of random rough surfaces in thin-film solar cells. *Commun. Comput. Phys.*, 25:1591–1612, 2019.

[9] Denis G. Baranov, Roman S. Savelev, Sergey V. Li, Alexander E. Krasnok, and Andrea Alù. Modifying magnetic dipole spontaneous emission with nanophotonic structures. *Laser & Photonics Reviews*, 11(3):1600268, April 2017.

[10] S. Basu, Z. M. Zhang, and C. J. Fu. Review of near-field thermal radiation and its application to energy conversion. *Int. J. Energy Res.*, 33(13):1203–1232, 2009.

[11] Elyas Bayati, Raphaël Pestourie, Shane Colburn, Zin Lin, Steven G. Johnson, and Arka Majumdar. Inverse designed extended depth of focus meta-optics for broadband imaging in the visible. *Nanophotonics*, 2021.

[12] Peter Bermel, Michael Ghebrebrhan, Walker Chan, Yi Xiang Yeng, Mohammad Araghchini, Rafif Hamam, Christopher H. Marton, Klavs F. Jensen, Marin Soljačić, John D. Joannopoulos, Steven G. Johnson, and Ivan Celanovic. Design and global optimization of high-efficiency thermophotovoltaic systems. *Optics Express*, 18:A314–A334, 2010.

[13] Jeff Bezanson, Alan Edelman, Stefan Karpinski, and Viral B. Shah. Julia: A fresh approach to numerical computing. *SIAM Rev.*, 59(1):65–98, 2017.

[14] A. D. Boardman. *Electromagnetic Surface Modes*. Wiley-Interscience, 1982.

[15] Jean-Daniel Boissonnat and Steve Oudot. Provably good sampling and meshing of surfaces. *Graphical Models*, 67(5):405–451, 2005. Solid Modeling and Applications.

[16] B. J. M. Brenny, T. Coenen, and A. Polman. Quantifying coherent and incoherent cathodoluminescence in semiconductors and metals. *J. Appl. Phys.*, 115(24):244307, 2014.

[17] Filippo Capolino, David R. Jackson, Donald R. Wilton, and Leopold B. Felsen. Comparison of methods for calculating the field excited by a dipole near a 2-d periodic material. *IEEE Trans. Antennas Propag.*, 55(6):1644–1655, 2007.

[18] V. P. Carey, G. Chen, C. Grigoropoulos, M. Kaviany, and A. Majumdar. A review of heat transfer physics. *Nanoscale Microscale Thermophys. Eng.*, 12(1):1–60, 2008.

[19] R. R. Chance, A. Prock, and R. Silbey. *Molecular Fluorescence and Energy Transfer Near Interfaces*, pages 1–65. John Wiley & Sons, Ltd, 2007.

[20] Wen Cho Chew. A new look at reciprocity and energy conservation theorems in electromagnetics. *IEEE Trans. Antennas Propag.*, 56(4):970–975, 2008.

[21] Edwin K. P. Chong and Stanislaw H. Zak. *An Introduction to Optimization*. Wiley-Interscience Publication, 2nd edition, 2001.

[22] Rasmus E. Christiansen, Jérôme Michon, Mohammed Benzaouia, Ole Sigmund, and Steven G. Johnson. Inverse design of nanoparticles for enhanced Raman scattering. *Opt. Express*, 28(4):4444–4462, 2020.

[23] Rasmus E Christiansen, Joakim Vester-Petersen, Søren Peder Madsen, and Ole Sigmund. A non-linear material interpolation for design of metallic nanoparticles using topology optimization. *Computer Methods in Applied Mechanics and Engineering*, 343:23–39, 2019.

[24] RICHARDC Compton, ROSSC McPhedran, Zorana Popovic, GABRIELM Rebeiz, Peter Tong, and D Rutledge. Bow-tie antennas on a dielectric half-space: Theory and experiment. *IEEE transactions on antennas and propagation*, 35(6):622–631, 1987.

[25] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms*. MIT, Cambridge, 3 edition, 2009.

[26] Giuseppe D'Aguanno, Nadia Mattiucci, Marco Centini, Michael Scalora, and Mark J. Bloemer. Electromagnetic density of modes for a finite-size three-dimensional structure. *Phys. Rev. E*, 69:057601, May 2004.

[27] T.A. Davis. *Direct Methods for Sparse Linear Systems*. SIAM, 2006.

[28] Alejandro R. Diaz and Ole Sigmund. A topology optimization method for design of negative permeability metamaterials. *Struct Multidisc Optim*, 41:163–177, 2010.

[29] Ahmet Cemal Durgun, Constantine A Balanis, Craig R Birtcher, and David R Allee. Design, simulation, fabrication and testing of flexible bow-tie antennas. *IEEE Transactions on Antennas and Propagation*, 59(12):4425–4435, 2011.

[30] John Earman. Curie's principle and spontaneous symmetry breaking. *International Studies in the Philosophy of Science*, 18(2-3):173–198, 2004.

[31] DJ Ehrlich and JY Tsao. A review of laser–microchemical processing. *Journal of Vacuum Science & Technology B: Microelectronics Processing and Phenomena*, 1(4):969–984, 1983.

[32] Alexei A. Erchak, Daniel J. Ripin, Shanhui Fan, Peter Rakich, John D. Joannopoulos, Erich P. Ippen, Gale S. Petrich, and Leslie A. Kolodziejski. Enhanced coupling to vertical radiation using a two-dimensional photonic crystal in a semiconductor light-emitting diode. *Appl. Phys. Lett.*, 78(5):563–565, 2001.

[33] Ying Fang, Nak-Hyun Seong, and Dana D Dlott. Measurement of the distribution of site enhancements in surface-enhanced Raman scattering. *Science*, 321(5887):388–392, 2008.

[34] Robin L Garrell. Surface-enhanced Raman spectroscopy. *Analytical Chemistry*, 61(6):401A–411A, 1989.

[35] Rong-Chun Ge, Philip Trøst Kristensen, Jeff F Young, and Stephen Hughes. Quasinormal mode approach to modelling light-emission and propagation in nanoplasmonics. *New Journal of Physics*, 16(11):113048, 2014.

[36] Christophe Geuzaine and Jean-François Remacle. Gmsh: A 3-d finite element mesh generator with built-in pre- and post-processing facilities. *Int. J. Numer. Methods Eng.*, 79(11):1309–1331, 2009.

[37] Walton C. Gibson. *The Method of Moments in Electromagnetics*. CRC Press, Boca Raton, FL, 3rd edition, 2021.

[38] Tao Gong, Matthew R. Corrado, Ahmed R. Mahbub, Calum Shelden, and Jeremy N. Munday. Recent progress in engineering the casimir effect–applications to nanophotonics, nanomechanics, and chemistry. *Nanophotonics*, 10(1):523–536, 2021.

[39] Jean-Jacques Greffet, Patrick Bouchon, Giovanni Brucoli, and Fran çois Marquier. Light emission by nonequilibrium bodies: Local kirchhoff law. *Phys. Rev. X*, 8:021008, 2018.

[40] Mats Gustafsson, Kurt Schab, Lukas Jelinek, and Miloslav Capek. Upper bounds on absorption and scattering. *New Journal of Physics*, 2020.

[41] Wolfgang Hackbusch. *Hierarchical Matrices: Algorithms and Analysis*. Springer, Berlin, 2015.

[42] Rafif E. Hamam, Aristeidis Karalis, J. D. Joannopoulos, and Marin Soljačić. Coupled-mode theory for general free-space resonant scattering of waves. *Phys. Rev. A*, 75:053801, May 2007.

[43] Alec M. Hammond, Ardavan Oskooi, Steven G. Johnson, and Stephen E. Ralph. Photonic topology optimization with semiconductor-foundry design-rule constraints. *Opt. Express*, 29:23916–23938, 2021.

[44] Encai Hao and George C Schatz. Electromagnetic fields around silver nanoparticles and dimers. *The Journal of chemical physics*, 120(1):357–366, 2004.

[45] Roger F. Harrington. *Time-Harmonic Electromagnetic Fields*. Wiley, 2nd edition, 2001.

[46] Matthias Heinkenschloss and Luis N. Vicente. Analysis of inexact trust-region sqp algorithms. *SIAM Journal on Optimization*, 12(2):283–302, 2002.

[47] M.F. Hutchinson. A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines. *Commun. Stat. B: Simul. Comput.*, 18(3):1059–1076, 1989.

[48] Mithat Idemen. Confluent tip singularity of the electromagnetic field at the apex of a material cone. *Wave motion*, 38(3):251–277, 2003.

[49] Michael Innes. Don't unroll adjoint: Differentiating ssa-form programs, 2018. Preprint at `https://arxiv.org/abs/1810.07951`.

[50] T. Inui, Y. Tanabe, and Y. Onodera. *Group Theory and Its Applications in Physics*. Springer, Berlin Heidelberg, 2012.

[51] John David Jackson. *Classical Electrodynamics*. Wiley, 3 edition, 1998.

[52] O. T. A. Janssen, A. J. H. Wachters, and H. P. Urbach. Efficient optimization method for the light extraction from periodically modulated leds using reciprocity. *Opt. Express*, 18(24):24522–24535, 2010.

[53] J.S. Jensen and O. Sigmund. Topology optimization for nano-photonics. *Laser Photonics Rev.*, 5(2):308–321, 2011.

[54] Jianming Jin. *The Finite Element Method in Electromagnetics*. Wiley-IEEE Press, 3rd edition, 2014.

[55] J.D. Joannopoulos, S.G. Johnson, J.N. Winn, and R.D. Meade. *Photonic Crystals: Modeling the Flow of Light - Second Edition*. Princeton University Press, 2008.

[56] Richard A. Johnson and Dean W. Wichern. *Applied Multivariate Statistics*. Pearson, 6th edition, 2018.

[57] S. G. Johnson, M. L. Povinelli, M. Soljačić, A. Karalis, S. Jacobs, and J. D. Joannopoulos. Roughness losses and volume-current methods in photonic-crystal waveguides. *Appl. Phys. B*, 81(2–3):283–293, 2005.

[58] Steven Johnson and John Joannopoulos. Block-iterative frequency-domain methods for Maxwell's equations in a planewave basis. *Opt. Express*, 8(3):173–190, 2001.

[59] Steven G. Johnson. The nlopt nonlinear-optimization package, 2021. `http://github.com/stevengj/nlopt`.

[60] Robin R Jones, David C Hooper, Liwu Zhang, Daniel Wolverson, and Ventsislav K Valev. Raman techniques: fundamentals and frontiers. *Nanoscale research letters*, 14(1):1–34, 2019.

[61] Karl Joulain, Rémi Carminati, Jean-Philippe Mulet, and Jean-Jacques Greffet. Definition and measurement of the local density of electromagnetic states close to an interface. *Phys. Rev. B*, 68:245405, Dec 2003.

[62] Karl Joulain, Jean-Philippe Mulet, François Marquier, Rémi Carminati, and Jean-Jacques Greffet. Surface electromagnetic waves thermally excited: Radiative heat transfer, coherence properties and casimir forces revisited in the near field. *Surface Science Reports*, 57(3):59 – 112, 2005.

[63] T. A. Kelf, Y. Sugawara, R. M. Cole, J. J. Baumberg, M. E. Abdelsalam, S. Cintra, S. Mahajan, A. E. Russell, and P. N. Bartlett. Localized and delocalized plasmons in metallic nanovoids. *Phys. Rev. B*, 74:245415, Dec 2006.

[64] Kwang-Je Kim. An analysis of self-amplified spontaneous emission. *Nucl. Instrum.*, 250(1):396–403, 1986.

[65] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. https://arxiv.org/abs/1412.6980, 2014.

[66] Derek M. Kita, Jérôme Michon, Steven G. Johnson, and Juejun Hu. Are slot and sub-wavelength grating waveguides better than strip waveguides for sensing? *Optica*, 5:1046–1054, 2018.

[67] Katrin Kneipp, Martin Moskovits, and Harald Kneipp. *Surface-enhanced Raman scattering: physics and applications*, volume 103. Springer Science & Business Media, 2006.

[68] Andrew V. Knyazev. Toward the optimal preconditioned eigensolver: Locally optimal block preconditioned conjugate gradient method. *SIAM J. Sci. Comput.*, 23(2):517–541, 2001.

[69] E. Kokiopoulou, J. Chen, and Y. Saad. Trace optimization and eigenproblems in dimension reduction methods. *Numer. Linear Algebra Appl.*, 18(3):565–602, 2011.

[70] Ken Kreutz-Delgado. The complex gradient operator and the cr-calculus, 2009. Preprint at `https://arxiv.org/abs/0906.4835`.

[71] Zeyu Kuang and Owen D. Miller. Computational bounds to light–matter interactions via local conservation laws. *Physical Review Letters*, 125(26), December 2020.

[72] Philippe Lalanne, Wei Yan, Kevin Vynck, Christophe Sauvan, and Jean-Paul Hugonin. Light interaction with photonic and plasmonic resonances. *Laser & Photonics Reviews*, 12(5):1700113, 2018.

[73] C. Lanczos. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *J. Res. Natl. Bur. Stand.*, 45(4):255–282, 1950.

[74] L.D. Landau, E.M. Lifšic, E.M. Lifshitz, L. P, L.P. Pitaevskii, J.B. Sykes, and M.J. Kearsley. *Statistical Physics: Theory of the Condensed State*. Elsevier Science, 1980.

[75] Judith Langer, Dorleta Aberasturi, Javier Aizpurua, Ramón Alvarez-Puebla, Baptiste Auguie, Jeremy Baumberg, Guillermo Bazan, Steven Bell, Anja Boisen, Alexandre Brolo, Jaebum Choo, Dana Cialla-May, Volker Deckert, Laura Fabris, Karen Faulds, Javier Garcia de Abajo, Royston Goodacre, Duncan Graham, Amanda Haes, and Luis Liz-Marzán. Present and future of surface enhanced Raman scattering. *ACS Nano*, 14, 01 2020.

[76] P.D. Lax. *Linear Algebra and Its Applications*. Wiley, 2013.

[77] B. S. Lazarov and O. Sigmund. Filters in topology optimization based on Helmholtz-type differential equations. *Int. J. Numer. Methods Eng.*, 86(6):765–781, 2011.

[78] Eric C Le Ru, Evan Blackie, Matthias Meyer, and Pablo G Etchegoin. Surface enhanced Raman scattering enhancement factors: a comprehensive study. *The Journal of Physical Chemistry C*, 111(37):13794–13803, 2007.

[79] Eric C Le Ru and Pablo G Etchegoin. Rigorous justification of the $|E|^4$ enhancement factor in surface enhanced Raman spectroscopy. *Chemical Physics Letters*, 423(1-3):63–66, 2006.

[80] Ren-Cang Li. Rayleigh quotient based optimization methods for eigenvalue problems. In *Series in Contemporary Applied Mathematics*, pages 76–108. HEP, 2015.

[81] Xiangdong Liang and Steven G. Johnson. Formulation for scalable optimization of microcavities via the frequency-averaged local density of states. *Opt. Express*, 21(25):30812–30841, 2013.

[82] Ho-Chun Lin, Zeyu Wang, and Chia Wei Hsu. Full-wave solver for massively multi-channel optics using augmented partial factorization. *arXiv preprint arXiv:2205.07887*, 2022.

[83] Zin Lin, Xiangdong Liang, Marko Lončar, Steven G. Johnson, and Alejandro W. Rodriguez. Cavity-enhanced second harmonic generation via nonlinear-overlap optimization. *Optica*, 3:233–238, February 2016.

[84] Chiyan Luo, Arvind Narayanaswamy, Gang Chen, and J. D. Joannopoulos. Thermal radiation from photonic crystals: A direct calculation. *Phys. Rev. Lett.*, 93:213905–213908, 2004.

[85] Ivan Markovsky. *Low Rank Approximation*. Springer-Verlag, London, 2012.

[86] Olivier J. F. Martin and Nicolas B. Piller. Electromagnetic scattering in polarizable backgrounds. *Phys. Rev. E*, 58:3909–3915, Sep 1998.

[87] Azardokht Mazaheri, Hamid Reza Fallah, and Javad Zarbakhsh. Application of ldos and multipole expansion technique in optimization of photonic crystal designs. *Opt. Quant. Electron.*, 45:67–77, 2013.

[88] Jeffrey M McMahon, Anne-Isabelle Henry, Kristin L Wustholz, Michael J Natan, R Griffith Freeman, Richard P Van Duyne, and George C Schatz. Gold nanoparticle dimer plasmonics: finite element method calculations of the electromagnetic enhancement to surface-enhanced Raman spectroscopy. *Analytical and bioanalytical chemistry*, 394(7):1819–1825, 2009.

[89] A Merlen, F Lagugné-Labarthet, and E Harté. Surface-enhanced Raman and fluorescence spectroscopy of dye molecules deposited on nanostructured gold surfaces. *The Journal of Physical Chemistry C*, 114(30):12878–12884, 2010.

[90] Jérôme Michon, Mohammed Benzaouia, Wenjie Yao, Owen D. Miller, and Steven G. Johnson. Limits to surface-enhanced Raman scattering near arbitrary-shape scatterers. *Opt. Express*, 27(24):35189–35202, Nov 2019.

[91] O. D. Miller, C. W. Hsu, M. T. H. Reid, W. Qiu, B. G. DeLacy, J. D. Joannopoulos, M. Soljačić, and S. G. Johnson. Fundamental limits to extinction by metallic nanoparticles. *Phys. Rev. Lett.*, 112:123903, Mar 2014.

[92] Owen D. Miller. *Photonic Design: From Fundamental Solar Cell Physics to Computational Inverse Design*. PhD thesis, Berkeley, 2012.

[93] Owen D. Miller, Ognjen Ilic, Thomas Christensen, M. T. Homer Reid, Harry A. Atwater, John D. Joannopoulos, Marin Soljačić, and Steven G. Johnson. Limits to the optical response of graphene and two-dimensional materials. *Nano Letters*, 17(9):5408–5415, 2017.

[94] Owen D. Miller, Steven G. Johnson, and Alejandro W. Rodriguez. Shape-independent limits to near-field radiative heat transfer. *Physical Review Letters*, 115:204302, 2015.

[95] Owen D. Miller, Athanasios G. Polimeridis, M. T. Homer Reid, Chia Wei Hsu, Brendan G. DeLacy, John D. Joannopoulos, Marin Soljačić, and Steven G. Johnson. Fundamental limits to optical response in absorptive systems. *Opt. Express*, 24(4):3329–3364, Feb 2016.

[96] Peter Milonni. Semiclassical and quantum-electrodynamical approaches in non-relativistic radiation theory. *Phys. Rep.*, 25:1–81, 1976.

[97] Sean Molesky, Pengning Chao, and Alejandro W. Rodriguez. $\mathbb{T}$ operator bounds on electromagnetic power transfer: Application to far-field cross sections, 2020. https://arxiv.org/abs/2001.11531.

[98] Sean Molesky, Zin Lin, Alexander Y. Piggott, Weiliang Jin, Jelena Vucković, and Alejandro W. Rodriguez. Inverse design in nanophotonics. *Nat. Photon.*, 12(11):659–670, 2018.

[99] Sean Molesky, Prashanth S. Venkataram, Weiliang Jin, and Alejandro W. Rodriguez. Fundamental limits to radiative heat transfer: Theory. *Phys. Rev. B*, 101:035408, Jan 2020.

[100] Sean Molesky, Prashanth S. Venkataram, Weiliang Jin, and Alejandro W. Rodriguez. Fundamental limits to radiative heat transfer: Theory. *Phys. Rev. B*, 101:035408, Jan 2020.

[101] Ramis Movassagh and Alan Edelman. Isotropic entanglement. *arXiv preprint arXiv:1012.5039*, 2010.

[102] Almir Mutapcic, Stephen Boyd, Ardavan Farjadpour, Steven G. Johnson, and Yehuda Avniel. Robust design of slow-light tapers in periodic waveguides. *Engineering Optimization*, 41:365–384, April 2009.

[103] Susumu Noda and Masayuki Fujita. Photonic crystal efficiency boost. *Nature Photonics*, 3:129–130, 2009.

[104] L. Novotny and B. Hecht. *Principles of Nano-Optics*. Cambridge University, 2 edition, 2012.

[105] H.M. Nussenzveig. *Causality and Dispersion Relations*. Academic Press, New York, 1972.

[106] S. Osher and R. Fedkiw. *Level Set Methods and Dynamic Implicit Surfaces*. SpringerVerlag New York, 2003.

[107] Ardavan Oskooi and Steven G. Johnson. Distinguishing correct from incorrect PML proposals and a corrected unsplit PML for anisotropic, dispersive media. *J. Comput. Phys.*, 230:2369–2377, 2011.

[108] Ardavan Oskooi and Steven G. Johnson. Electromagnetic wave source conditions. In Allen Taflove, Ardavan Oskooi, and Steven G. Johnson, editors, *Advances in FDTD Computational Electrodynamics: Photonics and Nanotechnology*, chapter 4, pages 65–100. Artech, Boston, 2013.

[109] F. Palacios-Gomez, L. Lasdon, and M. Engquist. Nonlinear optimization by successive linear programming. *Management Science*, 28(10):1106–1120, 1982.

[110] E. D. Palik. *Handbook of Optical Constants of Solids*. Elsevier Science, 1998.

[111] Ying Pan, Rasmus E. Christiansen, Jérôme Michon, Juejun Hu, and Steven G. Johnson. Topology optimization of surface-enhanced Raman scattering substrates. *arXiv.org e-Print archive*, January 2021.

[112] Michael Patra. *On Quantum Optics of Random Media*. PhD thesis, University of Leiden, 2015.

[113] F. P. Payne and J. P. R. Lacey. A theoretical analysis of scattering loss from planar optical waveguides. *Opt. Quantum Electron.*, 26:977–986, 1994.

[114] Raphaël Pestourie, Wenjie Yao, Boubacar Kanté, and Steven G. Johnson. Efficient inverse design of large-area metasurfaces for incoherent light. *arXiv.org e-Print archive*, July 2022.

[115] K. B. Petersen and M. S. Pedersen. *The Matrix Cookbook*. Technical University of Denmark, 2012.

[116] A. Pick, A. Cerjan, D. Liu, A. W. Rodriguez, A. D. Stone, Y. D. Chong, and Steven G. Johnson. Ab-initio multimode linewidth theory for arbitrary inhomogeneous laser cavities. *Phys. Rev. A*, 91:063806, 2015.

[117] Roberto Pilot, Raffaella Signorini, Christian Durante, Laura Orian, Manjari Bhamidipati, and Laura Fabris. A review on surface-enhanced Raman scattering. *Biosensors*, 9(2):57, 2019.

[118] Brian D Piorek, Seung Joon Lee, Juan G Santiago, Martin Moskovits, San-joy Banerjee, and Carl D Meinhart. Free-surface microfluidic control of surface-enhanced Raman spectroscopy for the optimized detection of airborne molecules. *Proceedings of the National Academy of Sciences*, 104(48):18898–18901, 2007.

[119] Athanasios G. Polimeridis, M. T. H. Reid, Weiliang Jin, Steven G. Johnson, Jacob K. White, and Alejandro W. Rodriguez. Fluctuating volume-current formulation of electromagnetic fluctuations in inhomogeneous media: Incandescence and luminescence in arbitrary geometries. *Phys. Rev. B*, 92:134202, 2015.

[120] Lidong Qin, Shengli Zou, Can Xue, Ariel Atkinson, George C Schatz, and Chad A Mirkin. Designing, fabricating, and imaging Raman hot spots. *Proceedings of the National Academy of Sciences*, 103(36):13300–13303, 2006.

[121] Søren Raza, Sergey I Bozhevolnyi, Martijn Wubs, and N Asger Mortensen. Nonlocal optical response in metallic nanostructures. *Journal of Physics: Condensed Matter*, 27(18):183204, apr 2015.

[122] M. T. H. Reid and S. G. Johnson. Efficient computation of power, force, and torque in BEM scattering calculations. *IEEE Transactions on Antennas and Propagation*, 63(8):3588–3598, 2015.

[123] M. T. Homer Reid, O. D. Miller, A. G. Polimeridis, A. W. Rodriguez, E. M. Tomlinson, and S. G. Johnson. Photon torpedoes and Rytov pinwheels: Integral-equation modeling of non-equilibrium fluctuation-induced forces and torques on nanoparticles, 2017. Preprint at `https://arxiv.org/abs/1708.01985`.

[124] F. Reif. *Fundamentals of Statistical and Thermal Physics*. McGraw-Hill Series in Fundamentals of Physics, New York, 1965.

[125] J. Revels, M. Lubin, and T. Papamarkou. Forward-mode automatic differentiation in Julia, 2016. Preprint at `https://arxiv.org/abs/1607.07892`.

[126] Alejandro W. Rodriguez, Ognjen Ilic, Peter Bermel, Ivan Celanovic, John D. Joannopoulos, Marin Soljačić, and Steven G. Johnson. Frequency-selective near-field radiative heat transfer between photonic crystal slabs: A computational approach for arbitrary geometries and materials. *Phys. Rev. Lett.*, 107:114302, 2011.

[127] Alejandro W. Rodriguez, M. T. Homer Reid, and Steven G. Johnson. Fluctuating surface-current formulation of radiative heat transfer: Theory and applications. *Phys. Rev. B*, 88:054305, 2013.

[128] Alejandro Rodriguez-Wong. *Fluctuation-induced interactions and nonlinear nanophotonics*. PhD thesis, Massachusetts Institute of Technology, 2010.

[129] Lavinia Rogobete, Hannes Schniepp, Vahid Sandoghdar, and Carsten Henkel. Spontaneous emission in nanoscopic dielectric particles. *Opt. Lett.*, 28(19):1736–1738, 2003.

[130] Charles Roques-Carmes, Nicholas Rivera, Ali Ghorashi, Steven E. Kooi, Yi Yang, Zin Lin, Justin Beroz, Aviram Massuda, Jamison Sloan, Nicolas Romeo, Yang Yu, John D. Joannopoulos, Ido Kaminer, Steven G. Johnson, and Marin Soljačić. A framework for scintillation in nanophotonics. *Science*, 375(6583), February 2022.

[131] Zhichao Ruan and Shanhui Fan. Superscattering of light from subwavelength nanostructures. *Phys. Rev. Lett.*, 105:013901, Jun 2010.

[132] Timothy W. Saucer and Vanessa Sih. Optimizing nanophotonic cavity designs with the gravitational search algorithm. *Opt. Express*, 21(18):20831–20836, Sep 2013.

[133] Philipp-Immanuel Schneider, Xavier Garcia Santiago, Victor Soltwisch, Martin Hammerschmidt, Sven Burger, and Carsten Rockstuhl. Benchmarking five global optimization approaches for nano-optical shape optimization and parameter reconstruction. *ACS Photonics*, 6(11):2726–2733, September 2019.

[134] Hyungki Shim, Lingling Fan, Steven G. Johnson, and Owen D. Miller. Fundamental limits to near-field optical response over any bandwidth. *Phys. Rev. X*, 9:011043, Mar 2019.

[135] Vasile Sima. *Algorithms for Linear-Quadratic Optimization.* CRC Press, 1996.

[136] Allan W. Snyder and John D. Love. *Optical Waveguide Theory.* Springer, USA, 1983.

[137] Diego M Solís, José M Taboada, Fernando Obelleiro, Luis M Liz-Marzán, and F Javier García de Abajo. Optimization of nanoparticle-based sers substrates through large-scale realistic simulations. *ACS photonics*, 4(2):329–337, 2017.

[138] G. Strang. *Computational Science and Engineering.* Wellesley, 2007.

[139] Wonjoo Suh, Zheng Wang, and Shanhui Fan. Temporal coupled-mode theory and the presence of non-orthogonal modes in lossless multimode cavities. *IEEE Journal of Quantum Electronics*, 40(10):1511–1518, October 2004.

[140] K. Svanberg. A class of globally convergent optimization methods based on conservative convex separable approximations. *SIAM J. Optim.*, 12(2):555–573, 2002.

[141] The CGAL Project. *CGAL User and Reference Manual.* CGAL Editorial Board, 5.0.2 edition, 2020.

[142] D. A. Tortorelli and P. Michaleris. Design sensitivity analysis: Overview and review. *Inverse Probl. Eng.*, 1(1):71–105, 1994.

[143] Lloyd N. Trefethen and David Bau. *Numerical Linear Algebra.* SIAM, 1997.

[144] Lloyd N. Trefethen and J. A. C. Weideman. The exponentially convergent trapezoidal rule. *SIAM Rev.*, 56(3):385–458, 2014.

[145] Shashanka Ubaru, Jie Chen, and Yousef Saad. Fast estimation of $\mathrm{tr}(f(A))$ via stochastic Lanczos quadrature. *SIAM J. Matrix Anal. Appl.*, 38:1075–1099, 2017.

[146] N. P. van Dijk, K. Maute, M. Langelaar, and F. van Keulen. Level-set methods for structural topology optimization: a review. *Struct. Multidisc. Optim.*, 48:437–472, 2013.

[147] Nico van Dijk, Kurt Maute, Matthijs Langelaar, and Fred Keulen. Level-set methods for structural topology optimization: A review. *Struct. Multidiscipl. Optim.*, 48:437–472, 2013.

[148] Ernst Jan R. Vesseur, F. Javier García de Abajo, and Albert Polman. Broadband purcell enhancement in plasmonic ring cavities. *Phys. Rev. B*, 82:165419, Oct 2010.

[149] Jelena Vučković, Matthew Pelton, Axel Scherer, and Yoshihisa Yamamoto. Optimization of three-dimensional micropost microcavities for cavity quantum electrodynamics. *Phys. Rev. A*, 66:023808, Aug 2002.

[150] Fengwen Wang, Rasmus Ellebæk Christiansen, Yi Yu, Jesper Mørk, and Ole Sigmund. Maximizing the quality factor to mode volume ratio for ultra-small photonic crystal cavities. *Appl. Phys. Lett.*, 113(24):241101, 2018.

[151] Fengwen Wang, Boyan Stefanov Lazarov, and Ole Sigmund. On projection methods, convergence and robust formulations in topology optimization. *Struct. Multidiscipl. Optim.*, 43(6):767–784, 2010.

[152] F. Wijnands, J. B. Pendry, F. J. Garcia-Vidal, P. M. Bell, P. J. Roberts, and L. Martín Moreno. Green's functions for maxwell's equations: application to spontaneous emission. *Optical and Quantum Electronics*, 29:199–216, 1997.

[153] Emil Wolf. *Introduction to the Theory of Coherence and Polarization of Light.* Cambridge University Press, 2007.

[154] Hongxing Xu. Theoretical study of coated spherical metallic nanoparticles for single-molecule surface-enhanced spectroscopy. *Applied physics letters*, 85(24):5980–5982, 2004.

[155] Yong Xu, Reginald K. Lee, and Amnon Yariv. Quantum analysis and the classical analysis of spontaneous emission in a microcavity. *Phys. Rev. A*, 61:033807, Feb 2000.

[156] Shancheng Yang, Yue Wang, and Handong Sun. Advances and prospects for whispering gallery mode microcavities. *Adv. Opt. Mater.*, 3(9):1136–1162, 2015.

[157] Yang Yang, John M Callahan, Tong-Ho Kim, April S Brown, and Henry O Everitt. Ultraviolet nanoplasmonics: A demonstration of surface-enhanced Raman spectroscopy, fluorescence, and photodegradation using Gallium nanoparticles. *Nano letters*, 13(6):2837–2841, 2013.

[158] Wenjie Yao. Fundamental limits to local density of states in absorptive system. Master's thesis, Massachusetts Institute of Technology, 2020.

[159] Wenjie Yao, Mohammed Benzaouia, Owen D. Miller, and Steven G. Johnson. Approaching the upper limits of the local density of states via optimized metallic cavities. *Opt. Express*, 28:24185–24197, 2020.

[160] Zongfu Yu, Aaswath Raman, and Shanhui Fan. Fundamental limit of nanophotonic light trapping in solar cells. *PNAS*, 107(41):17491–17496, 2010.

[161] Junxi Zhang, Lide Zhang, and Wei Xu. Surface plasmon polaritons: physics and applications. *Journal of Physics D: Applied Physics*, 45(11):113001, 2012.

[162] Shengli Zou and George C Schatz. Silver nanoparticle array structures that produce giant enhancements in electromagnetic fields. *Chemical Physics Letters*, 403(1-3):62–67, 2005.