



Trace formulation for photonic inverse design with incoherent sources

Yao, Wenjie; Verdugo, Francesc; Christiansen, Rasmus E.; Johnson, Steven G.

Published in:
Structural and Multidisciplinary Optimization

Link to article, DOI:
[10.1007/s00158-022-03389-5](https://doi.org/10.1007/s00158-022-03389-5)

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Yao, W., Verdugo, F., Christiansen, R. E., & Johnson, S. G. (2022). Trace formulation for photonic inverse design with incoherent sources. *Structural and Multidisciplinary Optimization*, 65(11), Article 336.
<https://doi.org/10.1007/s00158-022-03389-5>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



Trace formulation for photonic inverse design with incoherent sources

Wenjie Yao¹ · Francesc Verdugo² · Rasmus E. Christiansen^{3,4} · Steven G. Johnson⁵

Received: 25 November 2021 / Revised: 7 July 2022 / Accepted: 31 August 2022 / Published online: 15 November 2022
© The Author(s) 2022

Abstract

Spatially incoherent light sources, such as spontaneously emitting atoms, naively require Maxwell's equations to be solved many times to obtain the total emission, which becomes computationally intractable in conjunction with large-scale optimization (inverse design). We present a trace formulation of incoherent emission that can be efficiently combined with inverse design, even for topology optimization over thousands of design degrees of freedom. Our formulation includes previous reciprocity-based approaches, limited to a few output channels (e.g., normal emission), as special cases but generalizes to a continuum of emission directions by exploiting the low-rank structure of emission problems. We present several examples of incoherent-emission topology optimization, including tailoring the geometry of fluorescent particles, a periodically emitting surface, and a structure emitting into a waveguide mode, as well as discussing future applications to problems such as Raman sensing and cathodoluminescence.

Keywords Incoherent emission · Topology optimization · Inverse design

1 Introduction

Incoherent emission (light emission from random current sources) arises in many problems in optics: spontaneous emission (fluorescence) (Milonni 1976; Kim 1986; Polimeridis et al. 2015), thermal emission in both far (Carey et al. 2008) and near (Basu et al. 2009; Rodriguez et al. 2013) fields, scintillation (Brenny et al. 2014; Roques-Carmes et al. 2021), Casimir and van der Waals forces (Gong et al. 2021), Raman scattering in fluid suspensions (Pilot et al. 2019), incoherent incident waves (Wolf 2007) (which

can be transformed to random sources via the equivalence principle Harrington 2001), and even scattering from surface roughness via a Born approximation (Johnson et al. 2005). However, accurate modeling of such spatially random sources can pose severe computational challenges, because a direct approach would involve averaging the results of many simulations over an ensemble of sources (Rodriguez et al. 2011; Luo et al. 2004; Bao et al. 2019); the statistics (correlation functions) of the *sources* are known, but the difficulty is converting this into statistics (e.g., average power) of the resulting *fields*. In the cases of fluorescence (Polimeridis et al. 2015), near-field thermal radiation (Rodriguez et al. 2013), and Casimir forces (Gong et al. 2021), for example, tractable methods for arbitrary geometries were only obtained recently. This challenge is compounded when one wishes to perform *inverse design* (Moesky et al. 2018)—large-scale optimization of emission over many geometric parameters, perhaps even over “every pixel” of a design region via topology optimization (TopOpt) (Jensen and Sigmund 2011)—because one must then repeat the computation 10–1000 s of times as the design evolves, e.g., to maximize spontaneous emission (Rogobete et al. 2003; Liang and Johnson 2013; Wang et al. 2018; Yao et al. 2020) or Raman emission (Christiansen et al. 2020) from a single molecule, much less a distribution of sources.

Responsible Editor: Shikui Chen

✉ Steven G. Johnson
stevenj@mit.edu

- ¹ Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA
- ² CIMNE, Centre Internacional de Mètodes Numèrics a l'Enginyeria, Castelldefels, Spain
- ³ NanoPhoton—Center for Nanophotonics, Technical University of Denmark, Kgs. Lyngby, Denmark
- ⁴ Department of Civil and Mechanical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark
- ⁵ Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, USA

In this paper, we present a unified framework for inverse design of incoherent emission, combining a *trace formulation* adapted from recent work (Rodriguez et al. 2013; Polimeridis et al. 2015; Reid et al. 2017) (Sect. 2) with a new algorithm to *simultaneously* optimize the geometry and evolve to an accurate estimate of the average emission/trace (Sect. 2.5). We apply this framework to perform density-based TopOpt (Jensen and Sigmund 2011) on several example problems in two dimensions: fluorescence from an optimized nanoparticle (Sect. 4.1), enhanced emission from a corrugated surface analogous to a light-emitting diode (Erchak et al. 2001) (Sect. 4.2), and optimized emission into a waveguide (Sect. 4.3). 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. We show that this emission can be computed by a small number of “eigen-sources” of a Hermitian operator, which can be determined by a Rayleigh-quotient optimization (Li 2015) 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 (Sect. 2.3)—this unifies and generalizes known results based on Kirchhoff’s law of thermal radiation (Reif 1965; Greffet et al. 2018) or (more generally) reciprocity (Roques-Carnes et al. 2021; Janssen et al. 2010) for computing emission into a single planewave direction—but our alternative approach (Sect. 2.5) yields a small number of solves even for $K \rightarrow \infty$. The other well-known special case is that of a single-emitter location with a random orientation, which reduces to the local density of states (LDOS) via three Maxwell solves (Milonni 1976; Oskooi and Johnson 2013), and this appears as another low-rank special case in our formulation (Sect. 2.4). We believe that this computational framework will enable many future developments in the computational design of complex optical devices involving a wide variety of incoherent processes (Sect. 5).

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 (Jensen and Sigmund 2011). 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 (Lazarov and Sigmund 2011), which regularizes the problem by setting a minimum length scale on the design. (Additional manufacturing constraints can be imposed by well-known techniques Hammond et al. 2021, 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 formulation 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 (Molesky et al. 2018; Tortorelli and Michaleris 1994). Given the objective function and its derivatives, a variety of large-scale optimization algorithms are available; we use the CCSA/MMA method (Svanberg 2002). We employ a recent free/open-source finite-element method (FEM) package, Gridap.jl (Badia and Verdugo 2020), in the Julia language (Bezanson et al. 2017), 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 (Revels et al. 2016; Innes 2018).

2 Trace formulation

In this section, we first 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. 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. 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 as well as in the more general cases of a continuum of outputs.

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 (Jin 2014)

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

where $\epsilon(\mathbf{x}, \omega)$ 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 Jin 2014) into a linear equation:

$$\mathbf{A}\mathbf{u} = \mathbf{b}, \quad (2)$$

where \mathbf{A} is a matrix representing the Maxwell operator on the left hand of Eq. (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 (Joannopoulos et al. 2008).

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:

$$P = \mathbf{u}^\dagger \mathbf{O} \mathbf{u}, \quad (3)$$

where \dagger denotes the conjugate transpose (adjoint) and $\mathbf{O} = \mathbf{O}^\dagger$ is a Hermitian matrix/operator. In this paper, we are mainly concerned with computing emitted power P , which is constrained by the outgoing boundary conditions to be non-negative, in which case \mathbf{O} must furthermore be a *positive semidefinite* Hermitian matrix (i.e., non-negative eigenvalues) in the subspace of permissible \mathbf{u} , a property that will be useful in Sect. 2.5.

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 $\langle \dots \rangle$, of our quadratic form Eq. (3):

$$\langle P \rangle = \langle \mathbf{u}^\dagger \mathbf{O} \mathbf{u} \rangle = \langle \mathbf{b}^\dagger \mathbf{A}^{-\dagger} \mathbf{O} \mathbf{A}^{-1} \mathbf{b} \rangle, \quad (4)$$

where $\mathbf{A}^{-\dagger}$ denotes $(\mathbf{A}^{-1})^\dagger = (\mathbf{A}^\dagger)^{-1}$. 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 (Rodriguez et al. 2011; Luo et al. 2004), 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 (Rodriguez et al. 2013) and spontaneous emission (Polimeridis et al. 2015), 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 Reid

et al. 2017) is to write the scalar $\langle P \rangle$ as a 1×1 “matrix” trace and then employ the cyclic-shift property (Lax 2013) to group the \mathbf{b} terms together:

$$\langle P \rangle = \langle \mathbf{b}^\dagger \mathbf{A}^{-\dagger} \mathbf{O} \mathbf{A}^{-1} \mathbf{b} \rangle = \text{tr} \langle \mathbf{b}^\dagger \mathbf{A}^{-\dagger} \mathbf{O} \mathbf{A}^{-1} \mathbf{b} \rangle = \text{tr} [\mathbf{A}^{-\dagger} \mathbf{O} \mathbf{A}^{-1} \langle \mathbf{b} \mathbf{b}^\dagger \rangle]. \quad (5)$$

Here, the ensemble average is now confined to the $\langle \mathbf{b} \mathbf{b}^\dagger \rangle$ term, which is just the correlation matrix \mathbf{B} (Johnson and Wichern 2018) of the currents; such a matrix is positive semidefinite, so it can be factorized (Trefethen and Bau 1997) (for convenience below) as follows:

$$\langle \mathbf{b} \mathbf{b}^\dagger \rangle = \mathbf{B} = \mathbf{D} \mathbf{D}^\dagger, \quad (6)$$

for some known matrix \mathbf{D} . Further information about constructing the matrix \mathbf{B} or its factorization \mathbf{D} is given in “Appendix 1.” (For the case of finite-element discretizations, we show that \mathbf{B} is a sparse matrix that is straightforward to assemble and \mathbf{D} is, for example, a sparse Cholesky factor Davis 2006.) Algebraically, expressing our results in terms of \mathbf{D} below leads to convenient Hermitian matrices, but we show in “Appendix 2” that the final algorithms can easily employ \mathbf{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 (Landau et al. 1980), \mathbf{B} and \mathbf{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* depend on the discretization scheme as explained in “Appendix 1.” For instance, in the case of thermal and quantum fluctuations, the mean-square currents are given by the fluctuation–dissipation theorem (FDT; Landau et al. 1980), while for spontaneous emission, one can use the FDT with a “negative temperature” determined by the population inversion (Pick et al. 2015; Patra 2015).

Inserting Eq. (6) into Eq. (5), we obtain our objective as the trace of a deterministic Hermitian matrix \mathbf{H} (which is positive-semidefinite if \mathbf{O} is, as for power), given by

$$\langle P \rangle = \text{tr} \underbrace{[\mathbf{A}^{-1} \mathbf{D}]^\dagger \mathbf{O} [\mathbf{A}^{-1} \mathbf{D}]}_{\mathbf{H}}. \quad (7)$$

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. (7) is the computation of $\mathbf{A}^{-1} \mathbf{D}$. Recall that the $N \times N$ matrix \mathbf{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 $\mathbf{A}^{-1} (\mathbf{D} \mathbf{v})$ for any *single* right-hand side \mathbf{v} , typically because the matrix \mathbf{A} is *sparse* (mostly zero) as in finite-element

methods (Jin 2014), but computing the *whole* matrix $\mathbf{A}^{-1}\mathbf{D}$ corresponds to solving N right-hand sides. Equivalently, computing explicit (dense) matrix inverses \mathbf{A}^{-1} is typically prohibitively expensive (in both time and storage) for matrices arising in large physical systems (Davis 2006). Fortunately, a large number of “iterative” algorithms have been proposed for *estimating* matrix traces to any desired accuracy using relatively *few* matrix–vector products (Hutchinson 1989; Ubaru et al. 2017), and what remains is to find a method well-suited to inverse design.

2.3 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 equation (7) simplifies to only K scattering problems. This fact is a generalization of earlier results commonly derived from electromagnetic reciprocity (Chew 2008), such as the well-known Kirchhoff’s law of thermal radiation (reciprocity of emission and absorption) (Reif 1965) or analogous results for scintillation (Roques-Carmes et al. 2021). More generally, this simplification arises whenever the matrix \mathbf{O} in Eq. (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 “metals” optimization problems in which one is maximizing intensity at a focal spot (Bayati et al. 2021). In matrix notation for a discretized problem, this quantity corresponds to

$$P = \|\mathbf{E}(\mathbf{x}_1)\|^2 = \|\mathbf{e}_1^\dagger \mathbf{u}\|^2 = \mathbf{u}^\dagger \underbrace{\mathbf{e}_1 \mathbf{e}_1^\dagger}_{\mathbf{O}} \mathbf{u}, \quad (8)$$

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 (Lax 2013) matrix $\mathbf{O} = \mathbf{e}_1 \mathbf{e}_1^\dagger$, and the trace equation (7) simplifies to $\langle P \rangle = \mathbf{v}_1^* \mathbf{v}_1 = \|\mathbf{v}_1\|^2$, where

$$\mathbf{v}_1 = \mathbf{D}^\dagger \mathbf{A}^{-\dagger} \mathbf{e}_1 \quad (9)$$

and $\mathbf{A}^{-\dagger} \mathbf{e}_1$ 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 (Chew 2008).

Another important example where \mathbf{O} is low rank arises when the output P is the power in one (or more) orthogonal “wave channels” (Snyder and Love 1983), 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 $\|\mathbf{o}_1^\dagger \mathbf{u}\|^2$

(Snyder and Love 1983). Exactly as in the single-point case above, this corresponds to a rank-1 matrix $\mathbf{O} = \mathbf{o}_1 \mathbf{o}_1^\dagger$ 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) (Reif 1965; Greffet et al. 2018; Janssen et al. 2010). A similar technique was recently applied to optimize the average power emitted in the normal direction from a scintillation device (Roques-Carmes et al. 2021).

More generally, such cases correspond to an output quadratic form \mathbf{O} that takes a low-rank (Lax 2013) form:

$$\mathbf{O} = \sum_{i=1}^K \mathbf{o}_i \mathbf{o}_i^\dagger, \quad (10)$$

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. (10) into Eq. (7) and applying the cyclic-trace identity, we obtain

$$\begin{aligned} \langle P \rangle &= \sum_{i=1}^K \text{tr} \left[(\mathbf{A}^{-1} \mathbf{D})^\dagger \mathbf{o}_i \mathbf{o}_i^\dagger \mathbf{A}^{-1} \mathbf{D} \right] \\ &= \sum_{i=1}^K \mathbf{o}_i^\dagger \mathbf{A}^{-1} \mathbf{D} (\mathbf{A}^{-1} \mathbf{D})^\dagger \mathbf{o}_i \\ &= \sum_{i=1}^K \mathbf{v}_i^\dagger \mathbf{v}_i = \sum_{i=1}^K \|\mathbf{v}_i\|^2, \end{aligned} \quad (11)$$

where

$$\mathbf{v}_i = \mathbf{D}^\dagger \mathbf{A}^{-\dagger} \mathbf{o}_i \quad (12)$$

corresponds to a single “reciprocal” Maxwell solve $\mathbf{A}^{-\dagger} \mathbf{o}_i = (\mathbf{A}^{-T} \mathbf{o}_i^*)^*$ (a single scattering problem) for each i . (Electromagnetic reciprocity simply corresponds to the fact that $\mathbf{A}^T = \mathbf{A}$ for reciprocal materials Chew 2008.) Hence, the full trace—the average emission into K channels—can be computed with only K solves, and in many such cases $K = 1$.

2.4 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 (Milonni 1976; Oskooi and Johnson 2013). 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 (Liang and Johnson 2013; Wang et al. 2018; Yao et al. 2020). More generally, this situation corresponds to the correlation matrix \mathbf{B} being low-rank: if \mathbf{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 Oskooi and Johnson 2013) and β_i are uncorrelated random numbers with zero mean and unit mean-square. Then the correlation matrix $\mathbf{B} = \langle \mathbf{b}\mathbf{b}^\dagger \rangle$ is simply the rank- K matrix $\mathbf{B} = \sum_i \mathbf{b}_i \mathbf{b}_i^\dagger$. In this case, the trace simplifies to

$$\text{tr}\mathbf{H} = \sum_{i=1}^K \mathbf{u}_i^\dagger \mathbf{O} \mathbf{u}_i, \quad (13)$$

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

2.5 Trace computation: many output channels

In general, neither the matrix \mathbf{O} nor the matrix \mathbf{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 $\mathbf{H} = (\mathbf{A}^{-1} \mathbf{D})^\dagger \mathbf{O} (\mathbf{A}^{-1} \mathbf{D})$ from Eq. (7) is itself typically *approximately low rank* (“numerically low rank” Markovsky 2012) even if \mathbf{O} is not: the trace, which is equal to the sum of the eigenvalues of \mathbf{H} (Lax 2013), is *dominated by a few of \mathbf{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 \mathbf{H} (which we illustrate with numerical examples in Sect. 4). First, on physical grounds, emission enhancement arises due to resonances (via the Purcell effect) (Agio and Cano 2013), 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 density of states (Yu et al. 2010). The traditional definition of resonant modes corresponds to poles of \mathbf{A}^{-1} at complex resonant frequencies, which are (linear or nonlinear) eigenvalues ω satisfying $\det \mathbf{A}(\omega) = 0$ (Nussenzveig 1972); analogously, Eq. (7) decomposes the total power into a sum of eigenvalues corresponding to “resonant current” sources which diagonalize \mathbf{H} at a given frequency. More explicitly, if $\mathbf{A}^{-1} \mathbf{D}$ can be accurately approximated by the action of K resonances of \mathbf{A} (a quasinormal mode expansion Lalanne et al. 2018; Ge et al. 2014), so that \mathbf{A}^{-1} can be replaced by a rank- K matrix, it follows that \mathbf{H} is also approximately rank $\leq K$

(since it is a product of rank-deficient matrices Lax 2013). Moreover, geometric optimization to maximize the emitted power modifies the structure to further enhance one or more resonances (Liang and Johnson 2013), and we observe that this sometimes increases the concentration of the trace into a few eigenvalues of \mathbf{H} ; that is, optimized structures tend to be even lower rank. Second, in a more general mathematical sense, the matrix \mathbf{H} is built from *off-diagonal blocks* of the Green’s function matrix \mathbf{A}^{-1} , connecting sources (at the support of \mathbf{D}) to emitted power at some other location (the support of \mathbf{O} , e.g., where the Poynting flux is computed), and off-diagonal blocks of Green’s functions are known to be approximately low-rank (Hackbusch 2015). This is closely related to fast methods for integral equations, such as the fast-multipole method and others (Gibson 2021); essentially, far fields mostly depend on low-order spatial moments of the near fields/currents.

If $\text{tr}\mathbf{H}$ is dominated by $K \ll N$ the largest eigenvalues of the $N \times N$ matrix \mathbf{H} , then one merely needs a numerical algorithm to compute the K *extremal* (largest magnitude) eigenvalues using only a sequence matrix–vector products $\mathbf{H}\mathbf{v}$ (corresponding to individual scattering problems). Fortunately, there are many such algorithms, especially for Hermitian \mathbf{H} (Lanczos 1950; Knyazev 2001), 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 (Li 2015; Johnson and Joannopoulos 2001; Knyazev 2001; Kokiopoulou et al. 2011), and for positive semidefinite \mathbf{H} (= positive semidefinite \mathbf{O}) this sum is a lower bound on the trace (Kokiopoulou et al. 2011):

$$\text{tr}\mathbf{H} \geq \max_{\mathbf{V} \in \mathbb{C}^{N \times K}} \text{tr}[(\mathbf{A}^{-1} \mathbf{D} \mathbf{V})^\dagger \mathbf{O} (\mathbf{A}^{-1} \mathbf{D} \mathbf{V}) (\mathbf{V}^\dagger \mathbf{V})^{-1}], \quad (14)$$

where \mathbf{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 in Sect. 4.1 that the required K increases with the diameter of the emission region).

Computationally, one can maximize the right-hand side of Eq. (14) by some form of gradient ascent (Li 2015; Knyazev 2001), each step of which only requires the evaluation of $\mathbf{A}^{-1} \mathbf{D} \mathbf{V}$ for a $N \times K$ matrix \mathbf{V} . That is to say, one only needs K Maxwell solves at each step (instead of N for the full matrix \mathbf{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 \mathbf{V} to the geometry degrees of freedom and optimize over both \mathbf{V} and the geometry *simultaneously*. The full inverse-design problem with incoherent emission can now be bounded by a *single* optimization problem:

$$\langle P \rangle_{\text{optimum}} \geq \max_{\text{geometry}, \mathbf{V} \in \mathbb{C}^{N \times K}} \text{tr}[(\mathbf{A}^{-1} \mathbf{D} \mathbf{V})^\dagger \mathbf{O} (\mathbf{A}^{-1} \mathbf{D} \mathbf{V}) (\mathbf{V}^\dagger \mathbf{V})^{-1}], \quad (15)$$

where the geometric parameters (e.g., material densities Jensen and Sigmund 2011 or level sets van Dijk et al. 2013) only affect \mathbf{A} and (perhaps) \mathbf{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 (Molesky et al. 2018; Tortorelli and Michaleris 1994), whereas the gradient with respect to \mathbf{V} has a simple analytical formula (Johnson and Joannopoulos 2001) (“Appendix 3”), so a variety of gradient-based optimization algorithms (Chong and Zak 2001) can be applied to simultaneously evolve both \mathbf{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.

3 Topology-optimization formulation

In this section, we briefly review the density-based TopOpt formulation (Jensen and Sigmund 2011) that we employ for our example applications in Sect. 4. 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. (15), using a finite-element method (FEM) on a triangular mesh (Badia and Verdugo 2020; Jin 2014), 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 (Svanberg 2002).

Given a density $\rho(\mathbf{x}) \in [0, 1]$, one should first regularize the optimization problem by setting a non-strict minimum length scale 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 $\tilde{\rho}$ (Jensen and Sigmund 2011). 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 (Lazarov and Sigmund 2011):

$$\begin{aligned} -r_f^2 \nabla^2 \tilde{\rho} + \tilde{\rho} &= \rho, \\ \frac{\partial \tilde{\rho}}{\partial \mathbf{n}} \Big|_{\partial \Omega_D} &= 0, \end{aligned} \quad (16)$$

where r_f is the length scale 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 (Lazarov and Sigmund 2011). (In addition to this filtering, it is possible to impose additional fabrication/length scale constraints, for example to comply with semiconductor-foundry design rules Hammond et al. 2021.)

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 (Wang et al. 2010):

$$\tilde{\tilde{\rho}} = \frac{\tanh(\beta\eta) + \tanh(\beta(\tilde{\rho} - \eta))}{\tanh(\beta\eta) + \tanh(\beta(1 - \eta))}, \quad (17)$$

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 (Christiansen and Sigmund 2021); here, we used $\beta = 5, 10, 20, 40, 80$, similar to previous authors (Christiansen et al. 2020).

Finally, one obtains a material, described by an electric relative permittivity (dielectric constant) $\epsilon(\mathbf{r})$ in Eq. (1), given by

$$\epsilon(\mathbf{r}) = [\epsilon_1 + (\epsilon_2 - \epsilon_1) \tilde{\tilde{\rho}}(\mathbf{r})] \left(1 + \frac{i}{2Q} \right), \quad (18)$$

where ϵ_1 is the background material (usually air, $\epsilon_1 = 1$) and ϵ_2 is the design material (we use dielectric of $\epsilon_2 = 12$ throughout this work).

Equation (18) includes an optional “artificial loss” term $\sim 1/Q$, which effectively smooths out resonances to have quality factors $\leq Q$ (fractional bandwidth $\geq 1/Q$) (Liang and Johnson 2013). 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 (Liang and Johnson 2013). 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 (Liang and Johnson 2013).

The details of the FEM discretization are described in “Appendix 3,” but it is essentially a standard triangular mesh with first-order Lagrange elements (Jin 2014) and perfectly matched layers (PMLs) for absorbing boundaries (Oskooi and Johnson 2011). 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 Sect. 2) 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 (Molesky et al. 2018; Tortorelli and Michaleris 1994) with some help from automation (Revels et al. 2016), and then these are automatically composed (“backpropagated”) by an automatic-differentiation (AD) system (Innes 2018). 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 (Molesky et al. 2018).

4 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 Sect. 4.1 with a general

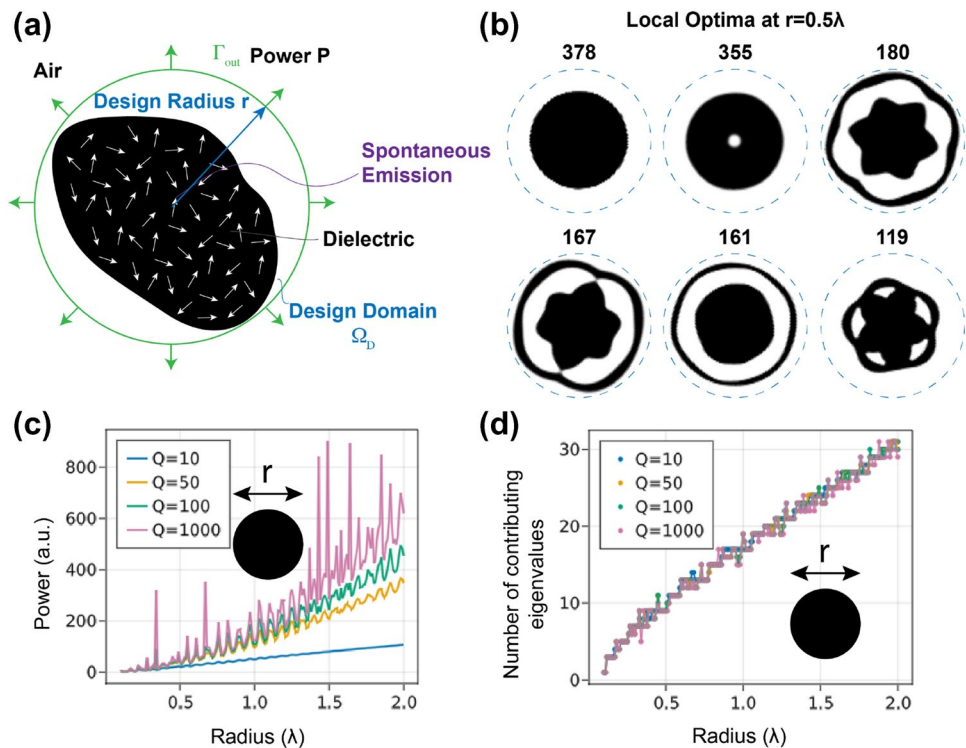
case where we are maximizing the total emitted power from many emitters distributed throughout a “fluorescent” dielectric material. Next, in Sect. 4.2, we study the enhanced emission from a corrugated surface, analogous to a light-emitting diode (Erchak et al. 2001), showing how the trace formulation can be applied to a periodic structure with aperiodic emitters. Both of these examples are based on the general algorithm from Sect. 2.5, which can handle emission into a continuum of possible angles. Finally, in Sect. 4.3, we apply the more specialized algorithm from Sect. 2.3 to optimizing emission from a fluorescent material into a single-mode waveguide. Since Maxwell’s equations are scale invariant (Joannopoulos et al. 2008), 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).

4.1 Fluorescent particle

In this example, illustrated in Fig. 1a, we optimize the shape/topology of a 2D fluorescent dielectric ($\epsilon = 12$) 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. Further computational details can be found in “Appendix C.1.”

Because this is a non-convex optimization problem, topology optimization can converge to different local optima from different initial geometries (Molesky et al. 2018). Figure 1b shows multiple local-optima geometries for a design radius

Fig. 1 **a** A 2D fluorescent particle (of dielectric $\epsilon = 12$) 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 unit. **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



$r = 0.5\lambda$ with filling ratio $R_f = 0.5$ and bandwidth quality factor $Q = 1000$ (artificial loss, from Sect. 3), 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 (Yao et al. 2020; Diaz and Sigmund 2010; Bermel et al. 2010), and while various heuristic strategies have been proposed to avoid poor local minima (Mutapcic et al. 2009; Aage and Egede Johansen 2017; Bermel et al. 2010; Schneider et al. 2019) beyond simply probing multiple random starting points, the only way to obtain rigorous guarantees is to derive theoretical upper bounds (Miller et al. 2016; Yao et al. 2020) as discussed further in Sect. 5 (purely numerical global search can generally provide practical guarantees only for very low-dimensional Maxwell optimization Azunre et al. 2019).

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 be increased 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. 1c, 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 (Yang et al. 2015).

The key assumption of our algorithm in Sect. 2.5 was that only a small number of eigenvalues would contribute to the trace, and this assumption clearly holds here. In Fig. 1d, 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) (Yu et al. 2010). However, we find that the scaling is nearly linear with the disk radius; the reason the simple DOS argument fails is that it does not 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.

4.2 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 (Erchak et al. 2001; Noda and Fujita 2009). As illustrated in Fig. 2a, the design domain consists of dielectric material ($\epsilon = 12$) 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 ($\epsilon = 2.25$) 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 Sect. 2.5. (Emission purely into the normal direction could be optimized much more efficiently using the methods of Sect. 2.3.) Further computational details can be found in “Appendix C.2”.

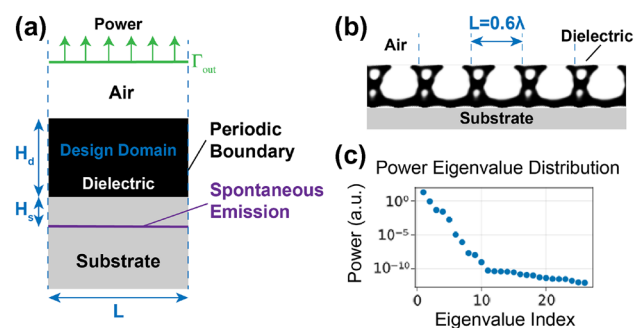


Fig. 2 **a** Unit cell of a 2D periodic emitting surface with period L . The design domain consists of dielectric material ($\epsilon = 12$) in air with thickness $H_d = 0.5\lambda$, the spontaneous-emission current sources are uniformly distributed on an horizontal line (purple line) inside a lower-index substrate ($\epsilon = 2.25$) 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

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” (Capolino et al. 2007). 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 ($\int_{-\pi/L}^{\pi/L} dk$) 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 Sect. 2. (Unit-cell calculations for different k values are completely independent and can be performed in parallel.) Further details of this formulation are described in “Appendix C.2.” (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 irreducible representations of the symmetry group Inui et al. 2012, thus, reducing the simulation domain even for asymmetrical random sources.)

The optimized structures for the design parameter $H_d = 0.5\lambda$, $H_s = 0.1\lambda$ are shown in Fig. 2b. 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. 2c: again, we observe that only the first few eigenvalues contribute significantly to the trace, as conjectured in Sect. 2.5.

4.3 Emission into a waveguide

This example considers a fluorescent dielectric ($\epsilon = 12$) medium in air, similar to Sect. 4.1, but in this case, we are maximizing the power coupled into a single-mode dielectric waveguide ($\epsilon = 12$, width $\lambda/2\sqrt{12}$) rather than into radiation (Fig. 3a). Since the output is a single channel (O is rank 1), this allows us to apply the method of Sect. 2.3 to perform only a single “reciprocal” Maxwell solve per optimization step. 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 of which 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 Liang and Johnson 2013; Wang et al. 2018). Further computational details can be found in “Appendix C.3.”

Figure 3b 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 Sect. 4.1, this area constraint was active at the optimum shown here. The corresponding averaged field intensity $\langle |H_z|^2 \rangle$ is displayed in Fig. 3c. 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 whole design domain is filled with $\epsilon = 12$ fluorescent material.

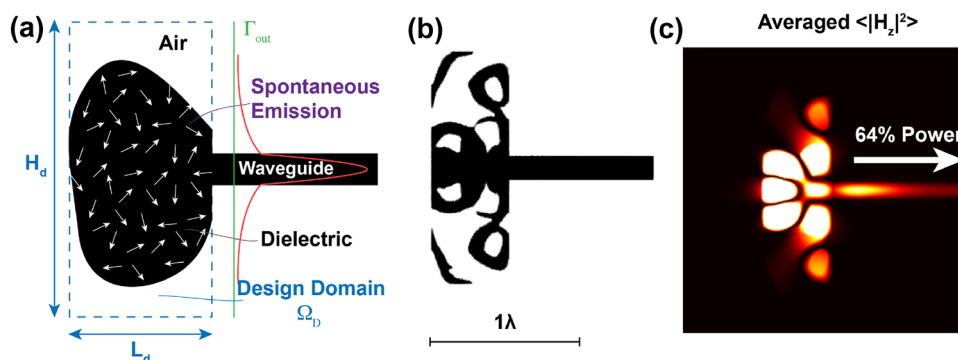


Fig. 3 **a** A 2D fluorescent dielectric ($\epsilon = 12$) medium coupling to a waveguide ($\epsilon = 12$, width $\lambda/2\sqrt{12}$). The design domain is of height $H_d = 1.5\lambda$ and width $L_d = 0.5\lambda$, the power coupled into a single-mode dielectric waveguide (mode overlap integral at Γ_{out}) is opti-

mized. **b** Optimized shape with a filling ratio $R_f = 0.5$. **c** Averaged field intensity $\langle |H_z|^2 \rangle$ distribution. About 64% of the power is coupled into the waveguide mode

5 Conclusion

We presented a trace formulation and accompanying algorithms for topology optimization of incoherent emitters, which unify and generalize earlier work, and in particular provide the first tractable optimization algorithms for the challenging case of many random emitters and many output channels. Looking forward, we believe that there are many potential applications of these ideas, as well as further algorithmic improvements and generalizations.

We are already preparing to 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 (Christiansen et al. 2020; Pan et al. 2021)—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 (Roques-Carnes et al. 2021). In contrast to spontaneous 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 (Wolf 2007), 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 (Johnson et al. 2005; Kita et al. 2018; Payne and Lacey 1994). 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 equation (15). When solving the eigenproblem alone (maximizing over \mathbf{V} to obtain extremal eigenvalues), it is well known that one can greatly improve upon straightforward gradient ascent by Krylov algorithms such as Arnoldi (Trefethen and Bau 1997) or LOBPCG (Knyazev 2001), 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 (Lin et al. 2022) 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 (Ubaru et al. 2017), one could further exploit

the fact that we are computing the trace of a function $f(\mathbf{A})$ of the Maxwell operator \mathbf{A} in order to relate the trace more efficiently to Krylov subspaces of \mathbf{A} . More generally, there are other applications where one is maximizing $\text{trf}(\mathbf{A}(p), p)$ 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 (Miller et al. 2015; Molesky et al. 2020) as well as to absorption (Kuang and Miller 2020; Miller et al. 2016) (related to emission via reciprocity), and many of these techniques should be adaptable to other forms of random emission.

Appendix 1: Correlation matrix

In this section, we show how to compute the correlation matrix \mathbf{B} corresponding to random current sources \mathbf{J} discretized in a finite-element basis. 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 $\nabla \times \mathbf{J}$ (Jin 2014). These two formulations lead to different \mathbf{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:

$$\langle \mathbf{J}(\mathbf{x})\mathbf{J}(\mathbf{x}')^\dagger \rangle = \mathbf{C}(\mathbf{x})\delta(\mathbf{x} - \mathbf{x}'), \quad (19)$$

where \mathbf{C} is a given 3×3 Hermitian positive-semidefinite correlation matrix. For example, in 2D with in-plane electric currents, as in the examples of Sect. 4, one has

$$\mathbf{C} = \begin{pmatrix} J_0^2 & & \\ & J_0^2 & \\ & & 0 \end{pmatrix}, \quad (20)$$

where $J_0^2(\mathbf{x})$ is the mean-square current at \mathbf{x} . For isotropic random currents, $\mathbf{C} = J_0^2 \mathbf{I}$ where \mathbf{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 $\hat{\mathbf{v}}_n \hat{\mathbf{z}}$ with scalar Lagrange elements \hat{v}_n in 2D for z -polarized fields) (Jin 2014). That is, the components of \mathbf{b} are

$$b_n = \int \hat{\mathbf{v}}_n \cdot (\text{source current}) \, d\Omega. \quad (21)$$

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

$$\begin{aligned} B_{mn} &= \langle b_m b_n^* \rangle, \\ &= \left\langle \iint \hat{\mathbf{v}}_m(\mathbf{x})^T \mathbf{J}(\mathbf{x}) \mathbf{J}(\mathbf{x}')^\dagger \hat{\mathbf{v}}_n(\mathbf{x}') d\Omega d\Omega' \right\rangle, \\ &= \iint \hat{\mathbf{v}}_m(\mathbf{x})^T \langle \mathbf{J}(\mathbf{x}) \mathbf{J}(\mathbf{x}')^\dagger \rangle \hat{\mathbf{v}}_n(\mathbf{x}') d\Omega d\Omega', \\ &= \int \hat{\mathbf{v}}_m^T \mathbf{C} \hat{\mathbf{v}}_n d\Omega. \end{aligned} \quad (22)$$

For localized basis functions (as in a finite-element method), this results in an extremely sparse matrix \mathbf{B} —it is zero if $\hat{\mathbf{v}}_m$ and $\hat{\mathbf{v}}_n$ do not overlap, or in regions where the mean-square current \mathbf{C} is zero. (If \mathbf{C} is the identity, \mathbf{B} is equal to the Gram matrix of the basis.) Note also that, by construction, \mathbf{B} is a Hermitian semidefinite matrix, so it has factorization $\mathbf{B} = \mathbf{D}\mathbf{D}^\dagger$, such as a Cholesky factorization (Trefethen and Bau 1997).

For a magnetic-field formulation, \mathbf{J} is replaced by $\nabla \times \mathbf{J}$ above, but we can simply integrate by parts (Joannopoulos et al. 2008) to move the $\nabla \times$ curl operation to act on the basis functions, yielding

$$B_{mn} = \langle b_m b_n^* \rangle = \int (\nabla \times \hat{\mathbf{v}}_m)^T \mathbf{C} (\nabla \times \hat{\mathbf{v}}_n) d\Omega. \quad (23)$$

Again, this yields a sparse Hermitian semidefinite matrix \mathbf{B} .

In the 2D examples of Sect. 4, we employed a magnetic-field formulation with an out-of-plane magnetic field $\mathbf{H} = H_z \hat{\mathbf{z}}$ and corresponding basis functions $\hat{\mathbf{v}}_n \hat{\mathbf{z}}$, along with in-plane current sources corresponding to Eq. (20). In this case, Eq. (23) simplifies to

$$B_{mn} = \int_{\Omega} J_0^2 (\nabla \hat{\mathbf{v}}_m \cdot \nabla \hat{\mathbf{v}}_n) d\Omega. \quad (24)$$

Appendix 2: Factorization-free trace formulation

Although it is conceptually attractive to use a trace formulation equation (7) in terms of the Hermitian matrix \mathbf{H} , this formulation required a factorization $\mathbf{B} = \mathbf{D}\mathbf{D}^\dagger$ of the correlation matrix \mathbf{B} . Computationally, it is desirable to avoid this factorization, especially if the current distribution (and hence \mathbf{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 problem equations (11) and (15) in terms of \mathbf{B} alone using a change of variables.

For the few-output-channel case in Sect. 2.3, one can simply start with Eq. (7) and rewrite it as $\langle P \rangle = \text{tr}[\mathbf{A}^{-\dagger} \mathbf{O} \mathbf{A}^{-1} \mathbf{B}]$, which for a low-rank \mathbf{O} simplifies, similar to Eq. (11), to

$$g(\rho) = \langle P \rangle = \sum_{i=1}^K \mathbf{u}_i^\dagger \mathbf{B} \mathbf{u}_i, \quad (25)$$

where $\mathbf{A}^\dagger \mathbf{u}_i = \mathbf{o}_i$, and we have defined the parameter ρ dependence (which can effect both \mathbf{A} and \mathbf{B}) as a function $g(\rho)$ for use in the adjoint formulation of “Appendix 3”.

For the many-channel case of Eq. (15), the key point is that we can choose \mathbf{V} to be orthogonal to the nullspace $N(\mathbf{D})$ of \mathbf{D} , as any nullspace component would contribute nothing to the trace ($\mathbf{D}\mathbf{V}$ projects it to zero). Equivalently, we can choose $\mathbf{V} = \mathbf{D}^\dagger \mathbf{W} \perp N(\mathbf{D})$ (Lax 2013) for any $N \times K$ matrix \mathbf{W} , and this change of variables yields a new optimization problem:

$$\begin{aligned} g(\rho, \mathbf{W}) &= \text{tr} \left[(\mathbf{A}^{-1} \mathbf{B} \mathbf{W})^\dagger \underbrace{\mathbf{O}}_{\mathbf{U}} (\mathbf{A}^{-1} \mathbf{B} \mathbf{W}) (\mathbf{W}^\dagger \mathbf{B} \mathbf{W})^{-1} \right] \\ &= \text{tr} [\mathbf{U}^\dagger \mathbf{O} \mathbf{U} (\mathbf{W}^\dagger \mathbf{B} \mathbf{W})^{-1}], \end{aligned} \quad (26)$$

where again we have defined the function $g(\rho, \mathbf{W})$ for the parameter and \mathbf{W} dependence, along with $\mathbf{U} = \mathbf{A}^{-1} \mathbf{B} \mathbf{W}$, for use in the adjoint formulation of “Appendix 3”.

Appendix 3: Numerical formulation

In this section, we provide details of the mathematical formulation and numerical implementation of the examples in Sect. 4, including the adjoint analysis.

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 :

$$\left[\nabla \times \frac{1}{\epsilon} \nabla \times - \left(\frac{\omega}{c} \right)^2 \mu \right] \mathbf{H}(\mathbf{x}) = \nabla \times \left[\frac{1}{\epsilon} \mathbf{J}(\mathbf{x}) \right]. \quad (27)$$

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 (Joannopoulos et al. 2008). In this case, Eq. (27) simplifies to a scalar Helmholtz equation:

$$\left[-\nabla \cdot \frac{1}{\epsilon} \nabla - \left(\frac{\omega}{c} \right)^2 \mu \right] H_z = \left(\nabla \times \left[\frac{1}{\epsilon} \mathbf{J}(\mathbf{x}) \right] \right) \cdot \hat{\mathbf{z}}. \quad (28)$$

Note that, for the correlation functions in the previous discussion, we simplified the right-hand side by absorbing the $1/\epsilon$ 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 $\nabla \rightarrow \Lambda \nabla$ in Eq. (28) (Oskooi and Johnson 2011; Jin 2014):

$$\left[-\Lambda \nabla \cdot \frac{1}{\varepsilon} \Lambda \nabla - \left(\frac{\omega}{c} \right)^2 \mu \right] H_z = \left(\nabla \times \left[\frac{1}{\varepsilon} \mathbf{J}(\mathbf{x}) \right] \right) \cdot \hat{\mathbf{z}}, \quad (29)$$

where

$$\Lambda = \begin{pmatrix} \frac{1}{1+i\sigma_x(\mathbf{x})/\omega} & & \\ & \frac{1}{1+i\sigma_y(\mathbf{x})/\omega} & \\ & & \frac{1}{1+i\sigma_z(\mathbf{x})/\omega} \end{pmatrix}. \quad (30)$$

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 (Oskooi and Johnson 2011), and we use a quadratic profile $\sigma_\ell(\mathbf{x}) = \sigma_0(x_{\text{PML}}/d_{\text{PML}})^2$ (where $x_{\text{PML}} \in [0, d_{\text{PML}}]$ is the distance inside the PML).

C.1: Fluorescent particle

For the problem of Sect. 4.1, the governing equation is exactly Eq. (29) with $\mu = 1$, whose weak form is (Jin 2014):

$$\begin{aligned} a(u, v) &= b(v), \\ a(u, v) &= \int_{\Omega} (\nabla \Lambda v \cdot \frac{1}{\varepsilon} \Lambda \nabla u - k_0^2 v u) d\Omega, \\ b(v) &= \int_{\Omega} v f d\Omega, \end{aligned} \quad (31)$$

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 \mathbf{A} and the source vector \mathbf{b} for the discretized Maxwell equation (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 (Jin 2014). The mesh was generated with Gmsh (Geuzaine and Remacle 2009), corresponding to a spatial resolution of roughly $\lambda/40$ in the air and $\lambda/80$ in the design region.

Notice that in Eq. (26), only \mathbf{U} (via \mathbf{A}) and \mathbf{B} (describing emission only in the dielectric) depend on the design parameters ρ . We have now the optimization problem as follows:

$$\begin{aligned} g(\rho, \mathbf{W}) &= \max_{\rho, \mathbf{W}} \text{tr}[\mathbf{U}(\rho)^\dagger \mathbf{O} \mathbf{U}(\rho) (\mathbf{W}^\dagger \mathbf{B}(\rho) \mathbf{W})^{-1}], \\ \mathbf{U}(\rho) &= \mathbf{A}(\rho)^{-1} \mathbf{B}(\rho) \mathbf{W}, \\ 0 &\leq \rho \leq 1, \\ \int \rho d\Omega_d &< \int R_f d\Omega_d, \end{aligned} \quad (32)$$

where R_f is the area-filling ratio.

Applying adjoint-method analysis (Molesky et al. 2018; Tortorelli and Michaleris 1994), we obtain the partial derivatives:

$$\begin{aligned} \frac{\partial g}{\partial \rho} &= -\text{tr} \left[\mathbf{U}^\dagger \mathbf{O} \mathbf{U} (\mathbf{W}^\dagger \mathbf{B} \mathbf{W})^{-1} \left(\mathbf{W}^\dagger \frac{\partial \mathbf{B}}{\partial \rho} \mathbf{W} \right) (\mathbf{W}^\dagger \mathbf{B} \mathbf{W})^{-1} \right] \\ &\quad - 2\text{Re} \left\{ \text{tr} \left[\mathbf{Z}^\dagger \left(\frac{\partial \mathbf{A}}{\partial \rho} \mathbf{U} - \frac{\partial \mathbf{B}}{\partial \rho} \mathbf{W} \right) \right] \right\}, \end{aligned} \quad (33)$$

where \mathbf{Z} is the result of an adjoint solve:

$$\mathbf{A}^\dagger \mathbf{Z} = \mathbf{O} \mathbf{U} (\mathbf{W}^\dagger \mathbf{B} \mathbf{W})^{-1}. \quad (34)$$

The partial derivative with respect to \mathbf{W} is simply obtained via matrix (Petersen and Pedersen 2012) CR calculus (Kreutz-Delgado 2009):

$$\frac{\partial g}{\partial \mathbf{W}} = [\mathbf{I} - \mathbf{B} \mathbf{W} (\mathbf{W}^\dagger \mathbf{B} \mathbf{W})^{-1} \mathbf{W}^\dagger] (\mathbf{A}^{-1} \mathbf{B})^\dagger \mathbf{O} \mathbf{U} (\mathbf{W}^\dagger \mathbf{B} \mathbf{W})^{-1}. \quad (35)$$

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. This CCSA update is implemented with NLOpt in Julia (Johnson 2021) 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.

C.2: Periodic emitting surface

For the problem of Sect. 4.2, we simulate a single unit cell with Bloch-periodic boundary conditions in x . Since Gridap only supports periodic boundary conditions in its current version, we make a change of variables $H_z \rightarrow H_z e^{ikx}$ so that

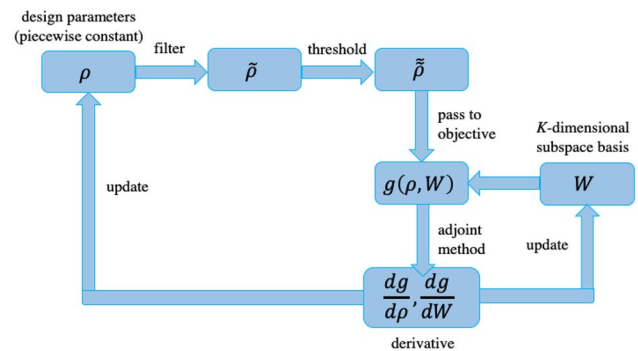


Fig. 4 Flowchart of the optimization steps for the fluorescent particle and periodic emitting surface examples

H_z is the periodic “Bloch envelope” function (Joannopoulos et al. 2008). In comparison to Eq. (28) in “Appendix C.1,” this corresponds to the transformation $\nabla \rightarrow \nabla + ik\hat{\mathbf{x}}$ (Joannopoulos et al. 2008):

$$\left[-(\nabla + ik\hat{\mathbf{x}}) \cdot \frac{1}{\varepsilon} (\nabla + ik\hat{\mathbf{x}}) - k_0^2 \right] H_z = f, \quad (36)$$

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

$$\begin{aligned} a(u, v) &= b(v), \\ a(u, v) &= \int_{\Omega} \left[(\nabla \Lambda - ik\hat{\mathbf{x}})v \cdot \frac{1}{\varepsilon} \cdot (\Lambda \nabla + ik\hat{\mathbf{x}})u - k_0^2 vu \right] d\Omega, \\ b(v) &= \int_{\Omega} v f d\Omega, \end{aligned} \quad (37)$$

where Λ is the diagonal PML “stretching” matrix equation (C12).

The objective (average power) is then constructed by a Brillouin-zone integration over the Bloch wavevector k (Capolino et al. 2007):

$$g(\rho) = \frac{L}{2\pi} \int_{-\pi/L}^{\pi/L} \text{tr} \left[(\mathbf{A}_k^{-1} \mathbf{D})^\dagger \mathbf{O} (\mathbf{A}_k^{-1} \mathbf{D}) \right] dk, \quad (38)$$

where L is the period of the unit cell and \mathbf{A}_k is assembled using Eq. (37). 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 (Trefethen and Weideman 2014); we used 100 k points in order to resolve sharp resonances.

Commuting the integral and the trace in Eq. (38), similarly to “Appendix 2” (noting that $\int \text{tr} = \text{tr} \int$), we obtain

$$\begin{aligned} g(\rho, \mathbf{W}) &= \max_{\rho, \mathbf{W}} \frac{L}{2\pi} \int_{-\pi/L}^{\pi/L} \text{tr} \left[\mathbf{U}_k(\rho)^\dagger \mathbf{O} \mathbf{U}_k(\rho) (\mathbf{W}^\dagger \mathbf{B}(\rho) \mathbf{W})^{-1} \right] dk, \\ \mathbf{U}_k(\rho) &= \mathbf{A}_k(\rho)^{-1} \mathbf{B}(\rho) \mathbf{W}, \\ 0 &\leq \rho \leq 1. \end{aligned} \quad (39)$$

The adjoint analysis for Eq. (39) is almost the same as in “Appendix C.1,” except for the additional integration over k . Also, it shares the same analysis workflow as in “Appendix C.1.”

C.3: Emission into a waveguide

For the problem of Sect. 4.3, the governing equation and the weak form are identical to “Appendix C.1.” The main difference is our objective function, which is now the power in a waveguide mode, computed via an overlap integral using mode orthogonality (Snyder and Love 1983), rather than a

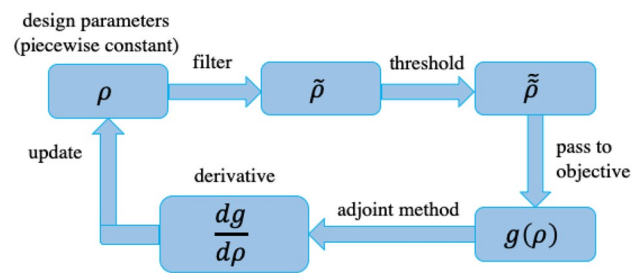


Fig. 5 Flowchart of the optimization steps for the emission into a waveguide example

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 (Snyder and Love 1983):

$$\alpha_i^* = \frac{\int \mathbf{e}_i \times \mathbf{H}^* \cdot d\mathbf{S}}{\int \mathbf{e}_i \times \mathbf{h}_i^* \cdot d\mathbf{S}} = \frac{\int e_{yi} H_z^* dy}{\int e_{yi} h_{zi}^* dy}, \quad (40)$$

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 Sect. 4.3, our objective is the power $|\alpha_0|^2$ in a single mode:

$$\langle P \rangle = |\alpha_0|^2 = \left| \frac{1}{N_0} \int e_{y0} H_z^* dy \right|^2, \quad (41)$$

where N_0 is the normalization (which can be omitted for optimization) from Eq. (40). 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. (41) becomes $\|\mathbf{o}^\dagger \mathbf{u}\|^2$ as in Eq. (10), where \mathbf{o} has components o_n given by the linear functional:

$$o_n = o(\hat{u}_n) = \frac{1}{N_0} \int e_{y0} \hat{u}_n dy. \quad (42)$$

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

$$\begin{aligned} g(\rho) &= \max_{\rho} [\mathbf{u}(\rho)^\dagger \mathbf{B}(\rho) \mathbf{u}(\rho)], \\ \mathbf{u}(\rho) &= \mathbf{A}(\rho)^{-1} \mathbf{o}, \\ 0 &\leq \rho \leq 1. \end{aligned} \quad (43)$$

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

$$\frac{dg}{d\rho} = \sum_{i=1}^K \left\{ \mathbf{u}_i^\dagger \frac{d\mathbf{B}}{d\rho} \mathbf{u}_i - 2\text{Re} \left[\mathbf{w}_i^\dagger \left(\frac{d\mathbf{A}^\dagger}{d\rho} \mathbf{u}_i \right) \right] \right\}, \quad (44)$$

where \mathbf{w}_i solves $\mathbf{A}\mathbf{w}_i = \mathbf{B}\mathbf{u}_i$ and \mathbf{u}_i solves the reciprocal problem $\mathbf{A}^\dagger \mathbf{u}_i = \mathbf{o}_i$ from Eq. (25). 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. 5.

Funding Open Access funding provided by the MIT Libraries. This work was supported in part by the U.S. Army Research Office through the Institute for Soldier Nanotechnologies under Award W911NF-13-D-0001, and by the PAPP Program of DARPA MTO under Award HR0011-20-90016. F. Verdugo acknowledges support from the Program Severo Ochoa Centre of Excellence (2019-2023) under the Grant CEX2018-000797-S funded by MCIN/AEI/10.13039/501100011033. R. E. Christiansen acknowledges support from the Danish National Research Foundation (Grant No. DNRF147 - NanoPhoton).

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Replication of results The code for Sect. 4 can be found at <https://github.com/WenjieYao/TraceFormula>.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aage N, Egede Johansen V (2017) Topology optimization of microwave waveguide filters. *Int J Numer Methods Eng* 112(3):283–300. <https://doi.org/10.1002/nme.5551>
- Agio M, Cano DM (2013) The Purcell factor of nanoresonators. *Nat Photonics* 7:674–675. <https://doi.org/10.1038/nphoton.2013.219>
- Azunre P, Jean J, Rotschild C, Bulovic V, Johnson S, Baldo M (2019) Guaranteed global optimization of thin-film optical systems. *N J Phys* 21:073050
- Badia S, Verdugo F (2020) Gridap: an extensible finite element toolbox in Julia. *J Open Source Softw* 5(52):2520. <https://doi.org/10.21105/joss.02520>
- Bao G, Cao Y, Lin J et al (2019) Computational optimal design of random rough surfaces in thin-film solar cells. *Commun Comput Phys* 25:1591–1612
- Basu S, Zhang ZM, Fu CJ (2009) Review of near-field thermal radiation and its application to energy conversion. *Int J Energy Res* 33(13):1203–1232. <https://doi.org/10.1002/er.1607>
- Bayati E, Pestourie R, Colburn S et al (2021) Inverse designed extended depth of focus meta-optics for broadband imaging in the visible. *Nanophotonics*. <https://doi.org/10.1515/nanoph-2021-0431>
- Bermel P, Ghebrebrhan M, Chan W et al (2010) Design and global optimization of high-efficiency thermophotovoltaic systems. *Opt Express* 18:A314–A334
- Bezanson J, Edelman A, Karpinski S et al (2017) Julia: a fresh approach to numerical computing. *SIAM Rev* 59(1):65–98. <https://doi.org/10.1137/141000671>
- Brenny BJM, Coenen T, Polman A (2014) Quantifying coherent and incoherent cathodoluminescence in semiconductors and metals. *J Appl Phys* 115(24):244307. <https://doi.org/10.1063/1.4885426>
- Capolino F, Jackson DR, Wilton DR et al (2007) 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. <https://doi.org/10.1109/TAP.2007.897348>
- Carey VP, Chen G, Grigoropoulos C et al (2008) A review of heat transfer physics. *Nanoscale Microscale Thermophys Eng* 12(1):1–60. <https://doi.org/10.1080/15567260801917520>
- Chew WC (2008) A new look at reciprocity and energy conservation theorems in electromagnetics. *IEEE Trans Antennas Propag* 56(4):970–975. <https://doi.org/10.1109/TAP.2008.919189>
- Chong EKP, Zak SH (2001) An introduction to optimization, 2nd edn. Wiley-Interscience Publication, New York
- Christiansen RE, Sigmund O (2021) Inverse design in photonics by topology optimization: tutorial. *J Opt Soc Am B* 38(2):496–509. <https://doi.org/10.1364/JOSAB.406048>
- Christiansen RE, Michon J, Benzaouia M et al (2020) Inverse design of nanoparticles for enhanced Raman scattering. *Opt Express* 28(4):4444–4462. <https://doi.org/10.1364/OE.28.004444>
- Davis T (2006) Direct methods for sparse linear systems. SIAM, Philadelphia
- Diaz AR, Sigmund O (2010) A topology optimization method for design of negative permeability metamaterials. *Struct Multidisc Optim* 41:163–177. <https://doi.org/10.1007/s00158-009-0416-y>
- Erchak AA, Ripin DJ, Fan S et al (2001) Enhanced coupling to vertical radiation using a two-dimensional photonic crystal in a semiconductor light-emitting diode. *Appl Phys Lett* 78(5):563–565. <https://doi.org/10.1063/1.1342048>
- Ge RC, Kristensen PT, Young JF et al (2014) Quasinormal mode approach to modelling light-emission and propagation in nanoplasmonics. *N J Phys* 16(11):113048. <https://doi.org/10.1088/1367-2630/16/11/113048>
- Geuzaine C, Remacle JF (2009) 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. <https://doi.org/10.1002/nme.2579>
- Gibson WC (2021) The method of moments in electromagnetics, 3rd edn. CRC Press, Boca Raton
- Gong T, Corrado MR, Mahbub AR et al (2021) Recent progress in engineering the Casimir effect—applications to nanophotonics, nanomechanics, and chemistry. *Nanophotonics* 10(1):523–536. <https://doi.org/10.1515/nanoph-2020-0425>
- Greffet JJ, Bouchon P, Brucoli G et al (2018) Light emission by nonequilibrium bodies: local Kirchhoff law. *Phys Rev X* 8(021):008
- Hackbusch W (2015) Hierarchical matrices: algorithms and analysis. Springer, Berlin
- Hammond AM, Oskooi A, Johnson SG et al (2021) Photonic topology optimization with semiconductor-foundry design-rule constraints. *Opt Express* 29:23916–23938. <https://doi.org/10.1364/OE.431188>
- Harrington RF (2001) Time-harmonic electromagnetic fields, 2nd edn. Wiley, New York
- Hutchinson M (1989) A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines. *Commun Stat B* 18(3):1059–1076. <https://doi.org/10.1080/03610918908812806>
- Innes M (2018) Don't unroll adjoint: differentiating SSA-form programs (Preprint). <https://arxiv.org/abs/1810.07951>

- Inui T, Tanabe Y, Onodera Y (2012) Group theory and its applications in physics. Springer, Berlin
- Janssen OTA, Wachters AJH, Urbach HP (2010) Efficient optimization method for the light extraction from periodically modulated LEDs using reciprocity. *Opt Express* 18(24):24522–24535. <https://doi.org/10.1364/OE.18.024522>
- Jensen J, Sigmund O (2011) Topology optimization for nano-photonics. *Laser Photonics Rev* 5(2):308–321. <https://doi.org/10.1002/lpor.201000014>
- Jin J (2014) The finite element method in electromagnetics, 3rd edn. Wiley-IEEE Press, New York
- Joannopoulos J, Johnson S, Winn J et al (2008) Photonic crystals: modeling the flow of light, 2nd edn. Princeton University Press, Princeton
- Johnson SG (2021) The NLOpt nonlinear-optimization package. <http://github.com/stevengj/nlopt>
- Johnson S, Joannopoulos J (2001) Block-iterative frequency-domain methods for Maxwell's equations in a planewave basis. *Opt Express* 8(3):173–190. <https://doi.org/10.1364/OE.8.000173>
- Johnson RA, Wichern DW (2018) Applied multivariate statistics, 6th edn. Pearson, Englewood Cliffs
- Johnson SG, Povinelli ML, Soljačić M et al (2005) Roughness losses and volume-current methods in photonic-crystal waveguides. *Appl Phys B* 81(2–3):283–293. <https://doi.org/10.1007/s00340-005-1823-4>
- Kim KJ (1986) An analysis of self-amplified spontaneous emission. *Nucl Instrum* 250(1):396–403. [https://doi.org/10.1016/0168-9002\(86\)90916-2](https://doi.org/10.1016/0168-9002(86)90916-2)
- Kita DM, Michon J, Johnson SG et al (2018) Are slot and sub-wave-length grating waveguides better than strip waveguides for sensing? *Optica* 5:1046–1054. <https://doi.org/10.1364/OPTICA.5.001046>
- Knyazev AV (2001) Toward the optimal preconditioned eigensolver: locally optimal block preconditioned conjugate gradient method. *SIAM J Sci Comput* 23(2):517–541. <https://doi.org/10.1137/S1064827500366124>
- Kokiopoulou E, Chen J, Saad Y (2011) Trace optimization and eigenproblems in dimension reduction methods. *Numer Linear Algebra Appl* 18(3):565–602. <https://doi.org/10.1002/nla.743>
- Kreutz-Delgado K (2009) The complex gradient operator and the CR-calculus (Preprint). <https://arxiv.org/abs/0906.4835>
- Kuang Z, Miller OD (2020) Computational bounds to light-matter interactions via local conservation laws. *Phys Rev Lett* 125(26):263607
- Lalanne P, Yan W, Vynck K et al (2018) Light interaction with photonic and plasmonic resonances. *Laser Photonics Rev* 12(5):1700113
- Lanczos C (1950) 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. <https://doi.org/10.6028/JRES.045.026>
- Landau L, Lifšic E, Lifšitz E et al (1980) Statistical physics: theory of the condensed state. Elsevier Science, Amsterdam
- Lax P (2013) Linear algebra and its applications. Wiley, Hoboken
- Lazarov BS, Sigmund O (2011) Filters in topology optimization based on Helmholtz-type differential equations. *Int J Numer Methods Eng* 86(6):765–781. <https://doi.org/10.1002/nme.3072>
- Li RC (2015) Rayleigh quotient based optimization methods for eigenvalue problems. Series in contemporary applied mathematics. HEP, pp 76–108
- Liang X, Johnson SG (2013) Formulation for scalable optimization of microcavities via the frequency-averaged local density of states. *Opt Express* 21(25):30812–30841. <https://doi.org/10.1364/OE.21.030812>
- Lin HC, Wang Z, Hsu CW (2022) Full-wave solver for massively multi-channel optics using augmented partial factorization. arXiv preprint. [arXiv:2205.07887](https://arxiv.org/abs/2205.07887)
- Luo C, Narayanaswamy A, Chen G et al (2004) Thermal radiation from photonic crystals: a direct calculation. *Phys Rev Lett* 93:213905–213908. <https://doi.org/10.1103/PhysRevLett.93.213905>
- Markovsky I (2012) Low rank approximation. Springer, London
- Miller OD, Johnson SG, Rodriguez AW (2015) Shape-independent limits to near-field radiative heat transfer. *Phys Rev Lett* 115(204):302
- Miller OD, Polimeridis AG, Reid MTH et al (2016) Fundamental limits to optical response in absorptive systems. *Opt Express* 24(4):3329–3364. <https://doi.org/10.1364/OE.24.003329>
- Milonni P (1976) Semiclassical and quantum-electrodynamical approaches in nonrelativistic radiation theory. *Phys Rep* 25:1–81. [https://doi.org/10.1016/0370-1573\(76\)90037-5](https://doi.org/10.1016/0370-1573(76)90037-5)
- Molesky S, Lin Z, Piggott AY et al (2018) Inverse design in nano-photonics. *Nat Photonics* 12(11):659–670. <https://doi.org/10.1038/s41566-018-0246-9>
- Molesky S, Venkataram PS, Jin W et al (2020) Fundamental limits to radiative heat transfer: theory. *Phys Rev B* 101(035):408. <https://doi.org/10.1103/PhysRevB.101.035408>
- Mutapcic A, Boyd S, Farjadpour A et al (2009) Robust design of slow-light tapers in periodic waveguides. *Eng Optim* 41:365–384
- Noda S, Fujita M (2009) Photonic crystal efficiency boost. *Nat Photonics* 3:129–130. <https://doi.org/10.1038/nphoton.2009.15>
- Nussenzveig H (1972) Causality and dispersion relations. Academic, New York
- Oskooi A, Johnson SG (2011) Distinguishing correct from incorrect PML proposals and a corrected unsplit PML for anisotropic, dispersive media. *J Comput Phys* 230:2369–2377. <https://doi.org/10.1016/j.jcp.2011.01.006>
- Oskooi A, Johnson SG (2013) Chap 4: electromagnetic wave source conditions. In: Taflov A, Oskooi A, Johnson SG (eds) *Advances in FDTD computational electrodynamics: photonics and nanotechnology*. Artech, Boston, pp 65–100
- Pan Y, Christiansen RE, Michon J et al (2021) Topology optimization of surface-enhanced Raman scattering substrates (Preprint). <https://arxiv.org/abs/2101.11352>
- Patra M (2015) On quantum optics of random media. PhD Thesis, University of Leiden
- Payne FP, Lacey JPR (1994) A theoretical analysis of scattering loss from planar optical waveguides. *Opt Quantum Electron* 26:977–986. <https://doi.org/10.1007/BF00708339>
- Petersen KB, Pedersen MS (2012) The matrix cookbook. Technical University of Denmark, Kgs. Lyngby
- Pick A, Cerjan A, Liu D et al (2015) Ab-initio multimode linewidth theory for arbitrary inhomogeneous laser cavities. *Phys Rev A* 91(063):806. <https://doi.org/10.1103/PhysRevA.91.063806>
- Pilot R, Signorini R, Durante C et al (2019) A review on surface-enhanced Raman scattering. *Biosensors* 9(2):57. <https://doi.org/10.3390/bios9020057>
- Polimeridis AG, Reid MTH, Jin W et al (2015) Fluctuating volume-current formulation of electromagnetic fluctuations in inhomogeneous media: incandescence and luminescence in arbitrary geometries. *Phys Rev B* 92(134):202. <https://doi.org/10.1103/PhysRevB.92.134202>
- Reid MTH, Miller OD, Polimeridis AG et al (2017) Photon torpedoes and Rytov pinwheels: integral-equation modeling of non-equilibrium fluctuation-induced forces and torques on nanoparticles (Preprint). <https://arxiv.org/abs/1708.01985>
- Reif F (1965) Fundamentals of statistical and thermal physics. McGraw-Hill series in fundamentals of physics. McGraw-Hill, New York

- Revels J, Lubin M, Papamarkou T (2016) Forward-mode automatic differentiation in Julia (Preprint). <https://arxiv.org/abs/1607.07892>
- Rodriguez AW, Ilic O, Bermel P et al (2011) Frequency-selective near-field radiative heat transfer between photonic crystal slabs: a computational approach for arbitrary geometries and materials. *Phys Rev Lett* 107(114):302. <https://doi.org/10.1103/PhysRevLett.107.114302>
- Rodriguez AW, Reid MTH, Johnson SG (2013) Fluctuating surface-current formulation of radiative heat transfer: theory and applications. *Phys Rev B* 88(054):305. <https://doi.org/10.1103/PhysRevB.88.054305>
- Rogobete L, Schniepp H, Sandoghdar V et al (2003) Spontaneous emission in nanoscopic dielectric particles. *Opt Lett* 28(19):1736–1738. <https://doi.org/10.1364/OL.28.001736>
- Roques-Carmes C, Rivera N, Ghorashi A et al (2021) A general framework for scintillation in nanophotonics (Preprint). <https://arxiv.org/abs/2110.11492>
- Schneider PI, Santiago XG, Soltwisch V et al (2019) Benchmarking five global optimization approaches for nano-optical shape optimization and parameter reconstruction. *ACS Photonics* 6(11):2726–2733
- Snyder AW, Love JD (1983) *Optical waveguide theory*. Springer, New York
- Svanberg K (2002) A class of globally convergent optimization methods based on conservative convex separable approximations. *SIAM J Optim* 12(2):555–573. <https://doi.org/10.1137/S1052623499362822>
- Tortorelli DA, Michaleris P (1994) Design sensitivity analysis: overview and review. *Inverse Probl Eng* 1(1):71–105. <https://doi.org/10.1080/174159794088027573>
- Trefethen LN, Bau D (1997) *Numerical linear algebra*. SIAM, Philadelphia
- Trefethen LN, Weideman JAC (2014) The exponentially convergent trapezoidal rule. *SIAM Rev* 56(3):385–458. <https://doi.org/10.1137/130932132>
- Ubaru S, Chen J, Saad Y (2017) Fast estimation of $\text{tr}(f(A))$ via stochastic Lanczos quadrature. *SIAM J Matrix Anal Appl* 38:1075–1099. <https://doi.org/10.1137/16M1104974>
- van Dijk N, Maute K, Langelaar M et al (2013) Level-set methods for structural topology optimization: a review. *Struct Multidisc Optim* 48:437–472. <https://doi.org/10.1007/s00158-013-0912-y>
- Wang F, Lazarov BS, Sigmund O (2010) On projection methods, convergence and robust formulations in topology optimization. *Struct Multidisc Optim* 43(6):767–784. <https://doi.org/10.1007/s00158-010-0602-y>
- Wang F, Christiansen RE, Yu Y et al (2018) Maximizing the quality factor to mode volume ratio for ultra-small photonic crystal cavities. *Appl Phys Lett* 113(24):241101. <https://doi.org/10.1063/1.5064468>
- Wolf E (2007) *Introduction to the theory of coherence and polarization of light*. Cambridge University Press, Cambridge
- Yang S, Wang Y, Sun H (2015) Advances and prospects for whispering gallery mode microcavities. *Adv Opt Mater* 3(9):1136–1162. <https://doi.org/10.1002/adom.201500232>
- Yao W, Benzaouia M, Miller OD et al (2020) Approaching the upper limits of the local density of states via optimized metallic cavities. *Opt Express* 28:24185–24197. <https://doi.org/10.1364/OE.397502>
- Yu Z, Raman A, Fan S (2010) Fundamental limit of nanophotonic light trapping in solar cells. *Proc Natl Acad Sci USA* 107(41):17491–17496. <https://doi.org/10.1073/pnas.1008296107>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.