



Validation and characterization of algorithms and software for photonics inverse design

MO CHEN,¹  RASMUS E. CHRISTIANSEN,^{2,3}  JONATHAN A. FAN,⁴  GÖKTAĞ IŞIKLAR,^{2,3} 
JIAQI JIANG,⁴  STEVEN G. JOHNSON,^{1,*}  WENCHAO MA,⁵ OWEN D. MILLER,⁶ 
ARDAVAN OSKOOL,⁷  MARTIN F. SCHUBERT,^{8,9}  FENGWEN WANG,³ 
IAN A. D. WILLIAMSON,⁸  WENJIN XUE,⁶ AND YOU ZHOU⁴ 

¹Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

²NanoPhoton, Center for Nanophotonics, Technical University of Denmark, Ørstedsgade 345A, 2800 Kongens Lyngby, Denmark

³Department of Civil and Mechanical Engineering, Solid Mechanics, Technical University of Denmark, DK-2800 Lyngby, Denmark

⁴Department of Electrical Engineering, Stanford University, Stanford, California 94305, USA

⁵Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

⁶Department of Applied Physics and Energy Sciences Institute, Yale University, New Haven, Connecticut 06511, USA

⁷Google, 1600 Amphitheatre Parkway, Mountain View, California 94043, USA

⁸X the moonshot factory, 100 Mayfield Avenue, Mountain View, California 94043, USA

⁹Current address: Meta, 1 Hacker Way, Menlo Park, California 94025, USA

*stevenj@math.mit.edu

Received 20 September 2023; revised 13 November 2023; accepted 9 December 2023; posted 21 December 2023; published 31 January 2024

In this work, we present a reproducible suite of test problems for large-scale optimization (“inverse design” and “topology optimization”) in photonics, where the prevalence of irregular, non-intuitive geometries can otherwise make it challenging to be confident that new algorithms and software are functioning as claimed. We include test problems that exercise a wide array of physical and mathematical features—far-field metalenses, 2d and 3d mode converters, resonant emission and focusing, and dispersion/eigenvalue engineering—and introduce an *a posteriori* lengthscales metric for comparing designs produced by disparate algorithms. For each problem, we incorporate cross-checks against multiple independent software packages and algorithms, and reproducible designs and their validation scripts are included. We believe that this suite should make it much easier to develop, validate, and gain trust in future inverse-design approaches and software. © 2024 Optica Publishing Group

<https://doi.org/10.1364/JOSAB.506412>

1. INTRODUCTION

Inverse design has become an increasingly important tool to develop non-intuitive/freeform solutions for challenging engineering problems in photonics [1]. Broadly speaking, inverse design consists of methods for large-scale optimization over vast numbers of geometric and material parameters—such as topology optimization [2], in which the material distribution at every point in space (“every pixel” in a discretized model) is a degree of freedom. New algorithms and software are continually being developed, but an obstacle to progress and wider adoption is a scarcity of accepted methods to rigorously validate new tools. This is because the result of inverse design is often a complex, non-intuitive structure for which no analytical solutions are available, and it can be unclear how to evaluate its performance. (In an increasing number of special cases, theoretical upper bounds are becoming available [3–5], but these bounds are not always tight: they may over-estimate the attainable performance.) It can be challenging to compare to previous inverse-designed structures in the literature (which are

often only published as an image in a figure), because different degrees of freedom and/or starting points may lead to entirely different local-optima designs [1]. For example, in many problems the results may depend sensitively on the imposed minimum lengthscales or other manufacturing constraints as the discretization model, but because there are many discretization schemes and distinct mathematical formulations of such constraints [6–10] it can be challenging to quantitatively compare results obtained by different methods. Existing validation approaches often compare solvers against simple analytical solutions or other software, check gradients (e.g., shape derivatives) against finite differences, verify optimality [Karush-Kuhn-Tucker (KKT)] conditions [11] at local optima, and compare optima with upper bounds, along with other good practices (for benchmarking, visualization, etc.) [12]. In addition to these useful checks, however, it would be valuable for the photonics inverse-design community to supplement basic tests with a widely accepted suite of optimization problems with reproducible performance (even if different algorithms obtain distinct

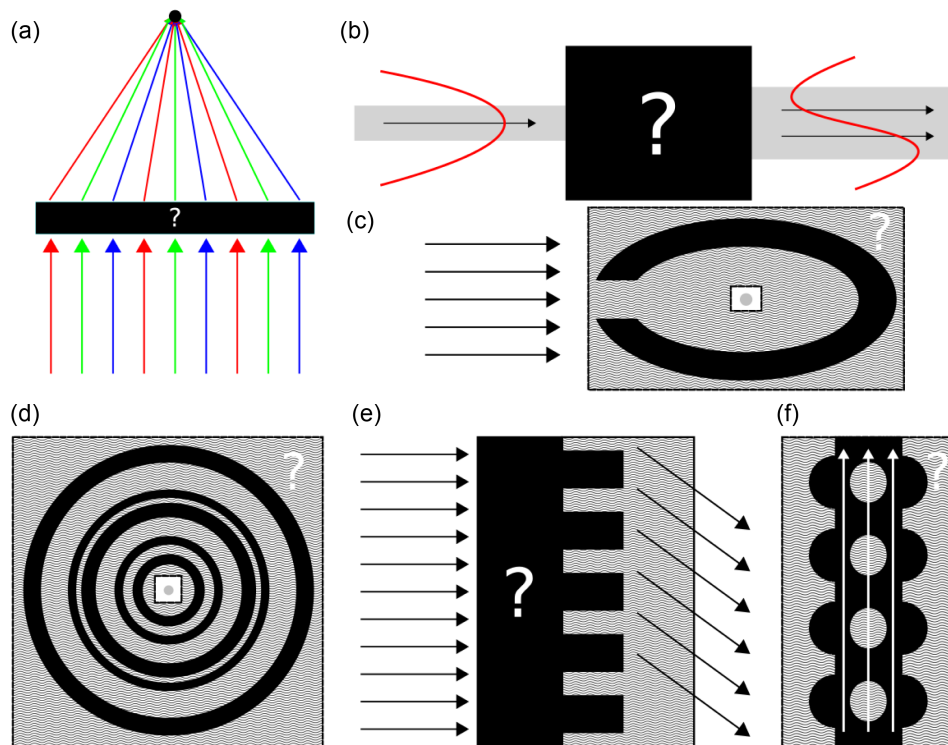


Fig. 1. Overview of the six testbed design problems for topology optimization used to validate various methods for photonics inverse design. The testbed problems are chosen to span a wide range of design objectives, dimensionality (2d and 3d), materials (dielectrics and metals), and design degrees of freedom. Several of the problems also involve imposing constraints on the minimum feature size. (a) RGB metalens. (b) Waveguide mode converter. (c) Metallic concentrator. (d) Purcell enhancement (LDOS). (e) Metagrating. (f) Slow-light waveguide.

local optima), for typical photonics-design problems that test a variety of algorithmic features.

To this end, our goal in this paper is to provide a test suite for photonics inverse-design tools: a set of carefully selected problems (Fig. 1), evaluated and characterized with multiple parameters—and in several cases checked with multiple independent software packages (thanks to a broad collaboration involving multiple research groups). Our problems are chosen to exercise typical important features of photonics inverse-design methods, such as minimax/worst-case optimization (e.g., Section 3.B), both single-wavelength (e.g., Sections 3.D–3.F) and broadband design (e.g., Section 3.C), modal decomposition (Section 3.C) or intensity-based objectives (e.g., Sections 3.B and 3.D), support for metals (Section 3.D) and dielectrics, both scattering (e.g., Sections 3.B, 3.C, and 3.F) and resonance (Section 3.E) or dispersion-relation (Section 3.G) design, near-to-far-field transformations (Section 3.B), as well as problems with fabrication/lengthscale constraints (e.g., Sections 3.B, 3.C, and 3.E). We include problems accessible to surface-integral-equation (e.g., boundary-element) methods (e.g., Sections 3.B and 3.D) as well as to partial-differential-equation (PDE, e.g., finite-element/finite-difference) methods. Each individual problem exercises only a few of these features, and the problems are small in scale—2d or low-diameter 3d (Section 3.F)—with the hope that any new algorithm or software should find at least one problem in the suite to be easily approachable. Dataset 1, Ref. [13] includes high-resolution data/images for exemplary

optimized designs. In order to assess the impact of fabrication/lengthscale constraints, we propose a simple *a posteriori* lengthscale-characterization (Section 2), so that in strongly lengthscale-dependent problems (e.g., Section 3.E) we can provide performance results as a function of lengthscale and results generated by different algorithms can be compared using the same metric. Free/open-source software is provided in Dataset 1, Ref. [13] [14,15] for the lengthscale metric and results, including scripts to validate the performance of complex inverse designs generated externally. For most non-trivial problems, one may not obtain *identical* results using different algorithms—the problems are typically non-convex (with many local optima), and algorithms may also involve different degrees of freedom and constraints—but we show that entirely different software should be able to obtain *comparable* performance results (even if the optimized geometries are different). Given an optimized structure, we also provide scripts (in Dataset 1, Ref. [13]) to validate its performance with free/open-source simulation software [16–19], or even to perform a new local optimization with the previous structure as a starting point (to check whether it can be further improved).

Since our test problems, along with most other practical optimization problems in photonics, are non-convex, optimization can only be expected to find local optima [1,2]. Global optimization algorithms, such as “multistart” algorithms in which one performs local optimization from random starting points [20], do not circumvent this issue because such algorithms cannot indicate when a global optimum has been reached (short of exhaustive search of the parameter space and related brute-force

methods, which are computationally infeasible except for a tiny number of parameters [3,21]). Different implementations of the same problem, even with the same starting guess and parameterizations, may converge to different local optima simply due to differences in the discretization or other details of the optimization algorithm, and an inflexible validation approach that required all users to find an identical local optimum would be onerous to use. However, distinct algorithms can still be usefully compared! Any local optimum is a lower bound to the global optimum, indicating *at least* how well one might hope to do, and for each problem we also report a range of local optima obtained by various starting points, giving some indication of the variance in expected results. For all of the test problems in this paper, we find that many local optima have performance comparable to the results shown, so that it should not be difficult to reproduce similar performance (if not the specific local minimum) in another solver. (Such diversity of good-performing local optima is often useful in inverse design, because it yields additional freedom to choose a good design according to other criteria such as ease of manufacturing; this is in stark contrast to the field of inverse scattering [22,23] in which a unique “ground truth” geometry is sought for imaging.) For some problems, an analytical upper bound is also available—e.g., mode conversion cannot exceed 100%, and new mathematical techniques are beginning to reveal upper bounds for light emission, focusing, and other effects [3–5]—so that the global optimum can be bracketed between optimization’s lower bound and analysis’ upper bound.

In the following sections, we first discuss a metric for evaluating the minimum lengthscales of a given design (Section 2) represented using a raster image and then present the problem formulations and validation results for a sequence of basic test problems (Section 3) for photonic inverse design (depicted schematically in Fig. 1). These problems consist of optimizing far-field three-wavelength (RGB) focusing by a 2d metalens [24] (Section 3.B), near-field single-wavelength focusing by a 2d metallic particle (Section 3.E), a 2d waveguide mode converter/demultiplexer [1,17,25–28] (Section 3.C), 3d diffraction-grating design [29–33] (Section 3.F), 2d resonance optimization via the local density of states (LDOS) [34–36] (Section 3.E), and 2d dispersion optimization for a photonic-crystal waveguide [37,38] (Section 3.G).

2. MEASURING MINIMUM LENGTHSCALES:

imageruler

A variety of techniques have been developed to *impose* minimum-lengthscale constraints during topology optimization [6–10]. The problem of *measuring* minimum lengthscales (or feature sizes), in order to characterize designs represented by raster images produced by different algorithms, is somewhat different, however. For example, lengthscale constraints must be differentiable in order to support optimization, whereas *a posteriori* measurement does not have this requirement; in some ways it is more analogous to design-rule checking (DRC) used in physical verification of semiconductor devices based on foundry specifications in process design kits (PDKs) [39,40]. DRCs, however, are not necessarily designed to extract a single lengthscale, and have several limitations for our application:

(1) their input geometry must be a polygon (a layer of a GDS file), (2) the KLayout DRC [39], in particular, does not reliably distinguish small lengthscales corresponding to sharp corners, and (3) the designs must be 2d.

Required attributes that we impose on our choice of lengthscale algorithm include the following.

- It must distinguish small lengthscales from mere discretization artifacts (e.g., single-pixel “bumps” arising in “staircased” interfaces).
- The lengthscale must be “resolution-invariant,” in that the same geometry sampled at different resolutions will yield nearly the same lengthscale (within an error imposed by the finite resolution).
- The lengthscale must be invariant under rotation/translation of the image (in the limit of infinite resolution).
- Sharp corners or cusps should be treated as arbitrarily small lengthscales at arbitrarily fine resolution (and more generally as a small lengthscale proportional to the pixel size).
- No lengthscale violations should occur at the outer borders of the image, since this depends on what “lies outside” the image (i.e., the boundary conditions, which can be chosen by suitably padding the image).
- The same algorithm should ideally be applicable to 1d (multilayer/grating), 2d, and even freeform 3d [41–46] design patterns.

Our resulting algorithm, described below and available as free/open-source software imageruler [14], is based on well-known morphological transformations, especially morphological opening and closing operations [8,47–50].

Given a binary image (density) ρ of “solid” (value = 1) and “void” (value = 0) regions like the one in Fig. 2(a), we want to determine separate minimum lengthscales for both solid and void (also called minimum “linewidth” and “linespacing” [51]), the locations of any violations in the image (features smaller than a given length d), as well as an overall minimum lengthscale equivalent to the smaller of the solid and void lengthscales. The central question is how one checks for violations given d , after which the minimum lengthscale can simply be defined as the largest d for which there is no violation at any $d' < d$. It suffices to explain how solid violations are determined, since void violations are equivalent to solid violations of the inverse image $\tilde{\rho} = 1 - \rho$.

We employ the mathematical tool of a morphological *opening* operation $\mathcal{O}_d(\rho)$ [47], which is defined as an *erosion* (shrinkage) of the solid regions followed by a *dilation* (expansion) of the remaining solid, in both cases using a kernel or structuring element given by a disc of diameter d (in order to preserve rotational invariance) as shown in Fig. 2(b), which corresponds intuitively to a round “probe” or “brush” [52]. (At diameters d comparable to the pixel size, where lengthscale measures are inherently less reliable, the “staircased” discretization of this disc kernel also becomes less symmetrical and more arbitrary, as shown in Supplement 1.) Any solid feature smaller than d is erased by the erosion operation and does not reappear during dilation, whereas larger features are restored to their original shapes. Hence, the *difference* $\rho - \mathcal{O}_d(\rho)$ is nonzero only at loci of lengthscale violations. However, in order to avoid treating

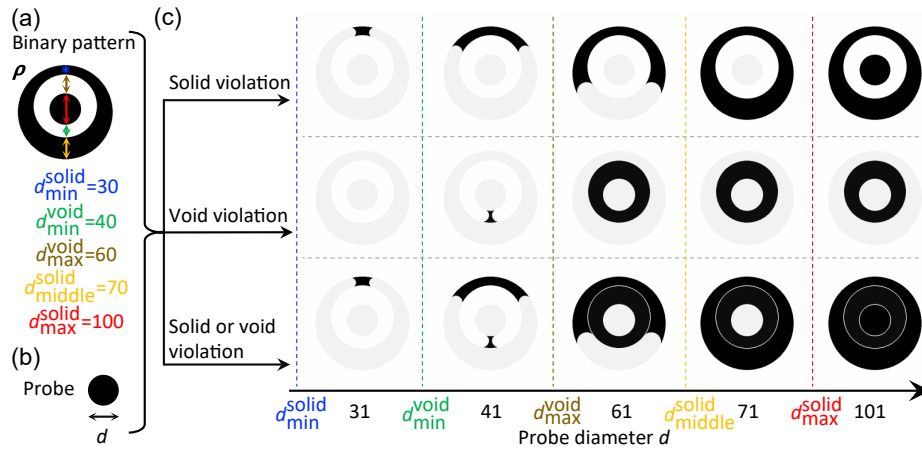


Fig. 2. Morphological method for estimating minimum lengthscales. (a) Example of a binary design pattern. The design pattern ρ is composed of nested discs with solid or void materials filled in between. All critical lengthscales are indicated by arrows and listed in the text with corresponding colors. In particular, the minimum lengthscales of solid and void regions are $d_{\min}^{\text{solid}} = 30$ and $d_{\min}^{\text{void}} = 40$. (b) Example of a probe. During the search for minimum lengthscales, the probe diameter d is subject to change. (c) Violation of various lengthscales. The first row demonstrates the violation of solid lengthscales, as described in Eq. (1). The second row demonstrates the same operation on the binary pattern with solid and void regions interchanged, as described in Eq. (2). The third row demonstrates the sum of the patterns in the two rows above, as described in Eq. (3). The original pattern is superimposed on each pattern of difference as a guide to the eye.

mere discretization artifacts as violations, one more important step is required: a pixel where $\rho - \mathcal{O}_d(\rho) \neq 0$ is only considered a violation if it lies in the *interior* of the original solid regions of ρ (not at the discretized solid/void interface), as defined more precisely in Supplement 1. This scheme is depicted for an example structure in the first row of Fig. 2(c) for varying brush diameters d : violations appear wherever d exceeds the minimum solid lengthscale d_{\min}^{solid} , defined as the smallest d for which a violation appears, i.e.,

$$[\rho - \mathcal{O}_d(\rho)] \cdot \mathcal{E}_\epsilon(\rho) \begin{cases} = 0, & d < d_{\min}^{\text{solid}}, \\ \neq 0, & d \geq d_{\min}^{\text{solid}}, \end{cases} \quad (1)$$

where $\mathcal{E}_\epsilon(\rho)$ denotes morphological erosion using a small kernel diameter ϵ (about 3 pixels), which generates the interior pixels of the solid regions in ρ . The value of d_{\min}^{solid} can be determined computationally by a binary search.

As mentioned above, the void lengthscales can be determined by an identical algorithm merely by inverting the original structure to $\bar{\rho} = 1 - \rho$: violations are nonzero pixels of $\bar{\rho} - \mathcal{O}_d(\bar{\rho})$ that lie in the interior of the solid regions in $\bar{\rho}$, corresponding to the interior of the original void regions. Mathematically, this is exactly equivalent to counting nonzero interior pixels of $\mathcal{C}_d(\rho) - \rho$, where $\mathcal{C}_d(\rho)$ is the morphological closing operation, consisting of dilation of the solid regions followed by erosion (instead of the reverse) [47]. This scheme is depicted in the second row of Fig. 2(c), in which void lengthscales violations appear as soon as d exceeds the minimum void lengthscales, i.e.,

$$[\bar{\rho} - \mathcal{O}_d(\bar{\rho})] \cdot \mathcal{E}_\epsilon(\bar{\rho}) = [\mathcal{C}_d(\rho) - \rho] \cdot \overline{\mathcal{D}_\epsilon(\rho)} \begin{cases} = 0, & d < d_{\min}^{\text{void}}, \\ \neq 0, & d \geq d_{\min}^{\text{void}}, \end{cases} \quad (2)$$

where $\mathcal{D}_\epsilon(\rho)$ denotes morphological dilation using a small kernel diameter ϵ , which satisfies $\mathcal{E}_\epsilon(\bar{\rho}) = \overline{\mathcal{D}_\epsilon(\rho)}$.

For many fabrication technologies, the constraints on minimum linespacing and linewidth differ, which is why it is useful to measure them separately [6–8,51]. However, to provide a simple metric for comparing different optimization algorithms on our test problems, we define a single overall minimum lengthscales given by the *minimum* of the solid and void lengthscales: $d_{\min} = \min(d_{\min}^{\text{solid}}, d_{\min}^{\text{void}})$. It turns out that there is another mathematically equivalent way to calculate this: because the solid and void regions are spatially disjoint, violation of either the solid or void minimum lengthscales is simply the *sum* of the two violations, which is equivalent to the violation calculated from the difference between morphological closing and opening, i.e.,

$$[\rho - \mathcal{O}_d(\rho)] \cdot \mathcal{E}_\epsilon(\rho) + [\bar{\rho} - \mathcal{O}_d(\bar{\rho})] \cdot \mathcal{E}_\epsilon(\bar{\rho}) = [\rho - \mathcal{O}_d(\rho)] \cdot \mathcal{E}_\epsilon(\rho) + [\mathcal{C}_d(\rho) - \rho] \cdot \overline{\mathcal{D}_\epsilon(\rho)} = [\mathcal{C}_d(\rho) - \mathcal{O}_d(\rho)] \cdot \left[\mathcal{E}_\epsilon(\rho) + \overline{\mathcal{D}_\epsilon(\rho)} \right] \begin{cases} = 0, & d < d_{\min}, \\ \neq 0, & d \geq d_{\min}. \end{cases} \quad (3)$$

This is depicted in the third row of Fig. 2(c), which shows violations for either solid or void, and is clearly the sum of the first two rows.

We have released a free/open-source Python-software package `imageruler` that implements these algorithms, with morphological transformations provided by the OpenCV software [53]. (The above algorithm is applied to 2d design patterns in the released version of our package because of the limitations of OpenCV, but we also implemented a version applicable to 3d designs. For 1d designs, the same algorithm could be used, but for simplicity the package just searches for the minimum lengths among all solid and/or void segments.) As explained above, `imageruler` effectively does not consider violations at the edge of the image; this is accomplished by padding the

image with solid when performing open operations and by void when performing close operations. (The user can instead pad the image with different boundary conditions as desired.) For the test problems below, we padded the design images according to boundary conditions defined in the problem. Further details of the implementation are described in [Supplement 1](#).

3. TEST PROBLEMS AND RESULTS

In the following sections, we describe our proposed test problems, along with exemplary results from a variety of photonics inverse-design software packages and algorithms. Our goal is to present these problems in a form that the problems could be implemented using many different algorithms, with many different parameterizations of the materials and design, besides those presented here. Although distinct algorithms and parameterizations may not converge to the same local minima, we typically find that comparable performance can be achieved, sometimes after exploring a modest number of local minima from different starting points. Given the specific local minima shown below (using data files supplied in [Dataset 1](#), Ref. [13]), of course, any numerical Maxwell solver should obtain the same results up to the limits of numerical resolution. We begin with a general overview of photonics inverse problems and the algorithms employed by our exemplary implementations.

A. Physical Models and Optimization Problems

For all test problems considered in the following, the physics is modeled using the classical Maxwell equations assuming linear, static, homogeneous, isotropic, non-dispersive, non-magnetic materials. Since the problem is linear time-invariant, the equations can ultimately be expressed in the frequency domain for time-harmonic fields ($\mathbf{E}e^{i\omega t}$, $\mathbf{H}e^{i\omega t}$) and excitations, even if in some cases we solve them in the time domain and obtain frequency-domain results by a discrete-time Fourier transform [17]) or via an equivalent integral-equation formulation [54,55]. The frequency-domain Maxwell equations can be written in terms of the electric fields, magnetic fields, or both:

$$\nabla \times \nabla \times \mathbf{E}(\mathbf{r}) - \frac{\omega^2}{c^2} \varepsilon(\mathbf{r}) \mathbf{E}(\mathbf{r}) = i\omega\mu_0 \mathbf{J}(\mathbf{r}), \quad \mathbf{r} \in \Omega, \quad (4)$$

$$\nabla \times \left(\frac{1}{\varepsilon(\mathbf{r})} \nabla \times \mathbf{H}(\mathbf{r}) \right) - \frac{\omega^2}{c^2} \mathbf{H}(\mathbf{r}) = -i\omega\varepsilon_0 \mathbf{K}(\mathbf{r}), \quad \mathbf{r} \in \Omega. \quad (5)$$

Here $\mathbf{E}(\mathbf{r})$ ($\mathbf{H}(\mathbf{r}) = \nabla \times \mathbf{E}/i\omega\mu_0$) denotes the spatially dependent electric (magnetic) field vector, ω the angular frequency, c the speed of light in vacuum, μ_0 (ε_0) the vacuum permeability (permittivity), ε the relative permittivity, $\Omega \subset \mathbb{R}^3$ the modeling domain, and \mathbf{J} (\mathbf{K}) an electric (magnetic) current density (the excitation source). These equations (or equivalent formulations thereof), along with a set of problem-appropriate boundary conditions, can be discretized in a variety of ways [54]; below, we employ either a finite-difference method, a finite-element method, a volume-integral-equation (VIE) method (which solves a corresponding integral-equation formulation [54–56]) in Section 3.B, or a modal-expansion method [57] in Section 3.F. Any frequency-domain discretization ultimately

results in a linear system of equations to be solved for a given material distribution and external forcing to obtain the electric and/or magnetic field in Ω . Of course, one also requires suitable boundary conditions; most of the test problems use outgoing/radiation boundary conditions (combined with periodic boundary conditions in Section 3.F), which were mainly implemented with perfectly matched layers (PML) [58–60], with the exception of integral-equation and modal-expansion methods where outgoing boundaries are analytically implicit in the method. Although a variety of units are popular for the Maxwell equations (including SI units and “dimensionless” units where $\varepsilon_0 = 1$ and $\mu_0 = 1$), our final optimization objective functions are expressed as normalized/non-dimensionalized values so that the units of a particular computational scheme become irrelevant.

Each inverse-design problem of this paper follows the typical form [1] of an optimization problem in which one parameterizes the geometry—the distribution $\varepsilon(\mathbf{r})$ of materials in space—as a function of some design parameters ρ that describe the structure in a design domain $\Omega_d \subseteq \Omega$, solving for the resulting electromagnetic fields $\mathbf{E}(\rho)$ and/or $\mathbf{H}(\rho)$, and then maximizes (or minimizes) a real-valued objective function $\Phi \in \mathbb{R}$ that depends on these fields (and possibly ρ), subject to $\mathcal{N}_j \geq 0$ constraints $c_j \leq 0$ (where $c_j \in \mathbb{R}$ is a function of ρ and/or the fields). That is, they are of the form

$$\begin{aligned} & \max_{\rho} \Phi(\mathbf{E}(\rho), \rho), \\ & \text{s.t. } c_j(\mathbf{E}(\rho), \rho) \leq 0, \text{ where } j \in \{1, \dots, \mathcal{N}_j\}. \end{aligned} \quad (6)$$

The electric (and/or magnetic) field in Eq. (6) is obtained by solving the corresponding (discretized) Maxwell equation given the material ε determined by the design parameters ρ , which can be thought of as an implicit equality constraint in the optimization problem. The constraints c_j (if any) can be used to represent manufacturing constraints, or can express other aspects of an optimization problem as discussed below.

In each of our test problems below, there will only be two options for materials (depending on the problem) at any point in Ω_d . There are many possible ways of parameterizing the arrangement of two materials in a given design domain Ω_d , from material density functions or level sets (which can describe any possible arrangement of two or more materials) to simple shape parameters (such as the diameters and centers of cylinders of material). To make (local) optimization tractable with large numbers of parameters, one typically requires ε to be parameterized *continuously* and *differentiably* in terms of continuous parameters ρ [1]. In our exemplary implementations of the test problems, we employ density-based topology optimization [2], in which ρ describes a function $\rho(\mathbf{r}) \in [0, 1]$ over the design domain $\mathbf{r} \in \Omega_d$, discretized in space via some basis or mesh (we used the same grid/mesh as the electromagnetic fields), where $\rho = 0$ or 1 refers conceptually to one material or the other. In practice, matters are somewhat more complicated: the design density ρ is first smoothed (low-pass filtered) to a function $\tilde{\rho}$ and then projected [61] using a continuation strategy to obtain a nearly binarized design $\hat{\rho}$. In our discretizations based on a uniform grid (finite difference and modal expansion), the smoothing $\rho \mapsto \tilde{\rho}$ simply employs convolution with a conic

filter, whereas in finite-element methods with an unstructured mesh we employ a damped-diffusion filter [62], in both cases with a problem-dependent smoothing radius r_f . The projection $\tilde{\rho} \mapsto \hat{\rho}$ is a smoothed step function:

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

$$\beta \in [1, \infty), \quad \eta \in [0, 1], \quad (7)$$

where $\eta = 0.5$ (except where noted below for robust optimization), and $\beta \in [1, \infty)$ is gradually increased during optimization to converge to a mostly binary design ($\hat{\rho} \approx 0$ or 1 almost everywhere).

Given the smoothed/projected design field $\hat{\rho}(\mathbf{r})$, the relative permittivity $\varepsilon(\mathbf{r})$ is then interpolated (either linearly in ε , linearly in ε^{-1} , or linearly in the refractive index $\sqrt{\varepsilon}$). The choice of interpolation depends in part on the nature of the particular test problem, and particular care is required for metallic design problems (Section 3.D) to ensure that intermediate densities $\hat{\rho} \in (0, 1)$ do not yield drastic unphysical behaviors [2,17,63,64].

In test problems with geometric lengthscale (manufacturing) constraints, we employ known constraints $c_j \leq 0$ of the form [6,51]

$$G_{\{s,v\}}(\hat{\rho}(x, y)) \leq 0, \quad (8)$$

where G_s and G_v enforce, respectively, the lengthscale constraints on the solid (ε_{\max}) and void (ε_{\min}) phases of the design. (Note that, in order to constrain the lengthscales at edges and corners of Ω_d , the design region is padded with the neighboring materials, and filtered through the padding. The design degrees of freedom are within the design region, but the padding and filtering lead to structures outside the design region in order to have smooth transition to the surrounding material.)

Another important use of constraint functions is for “maximin” objectives in which one is maximizing the *minimum* of $K > 1$ objective functions, i.e., $\Phi = \min_k \Phi_k$ for K functions $\Phi_k(\mathbf{E}(\rho), \rho)$, where $k \in \{1, \dots, K\}$. This arises in the metalens problem of Section 3.B, for example. Although such a Φ is non-differentiable where two Φ_k values cross, it can be transformed into a mathematically equivalent differentiable problem by an “epigraph” reformulation [11], which introduces a new design variable $t \in \mathbb{R}$ along with a set of K nonlinear constraints:

$$\max_{\rho} \left[\min_k \Phi_k \right] = \max_{\rho, t} t$$

$$\text{s.t. } t - \Phi_k(\mathbf{E}(\rho), \rho) \leq 0, \quad k \in \{1, \dots, K\}, \quad (9)$$

in addition to any other (e.g., lengthscale) constraints of the problem.

These optimization problems are solved using gradient-based algorithms to obtain a local minimum from a given starting ρ , relying on adjoint sensitivity analysis [2,65] to efficiently compute the gradients of the objective function and constraints with respect to the design variables ρ . Except where otherwise noted below, we used the CCSA/MMA optimization algorithm to obtain local optima [66].

Of course, there are many other ways to express the design degrees of freedom, as well as different algorithms for imposing constraints or for solving the optimization problem. Slight changes may result in a different local optimum, but in the problems below we find many local optima with comparable performance, and we expect that should continue to be true for any other algorithm with similar design freedom. Severely reducing the space of allowed designs, such as in the “1d” grating-design example of Section 3.F, is likely to reduce performance, but it is still informative to compare to reference designs with more freedom in order to quantify this degradation (and drastically worse performance may spur further investigation).

B. RGB Metalens

Inverse design of an aberration-free metalens [67] involving a single focus at multiple wavelengths is a well-studied design problem in photonics topology optimization [24,68–70], and the design of a metalens is thus chosen as the first test problem. One relevant design objective is to maximize the minimum intensity at the focal point for the RGB wavelengths. This is known as worst-case or maximin optimization [11]. Efficiently computing the fields at the focal point a large distance $\gg \lambda$ from the lens can involve a near-to-far-field transformation [71] or integral-equation methods [54–56].

As depicted in Fig. 3(a), in this test problem we look for the dielectric profile $\varepsilon(x, y) \in [\varepsilon_{\min}, \varepsilon_{\max}]$, “binarized” (almost) everywhere to ε_{\min} or ε_{\max} , in the design domain, Ω_d consisting of an $L_x = 10 \mu\text{m}$ by $L_y = 1 \mu\text{m}$ region that takes light at normal incidence at three wavelengths ($\lambda_1 = 450 \text{ nm}$, $\lambda_2 = 550 \text{ nm}$, $\lambda_3 = 650 \text{ nm}$) and focuses it at a spot, \mathbf{r}_0 , which is located $R = 2.4 \mu\text{m}$ above the center point in the design domain. Specifically, the objective is to maximize the worst-case electric-field intensity Φ over the targeted wavelengths at the focal spot:

$$\Phi = \left[\min_{\lambda_j} |\mathbf{E}_j(\mathbf{r}_0)|^2 \right], \quad (10)$$

where $\mathbf{E}_j(\mathbf{r}_0)$ is the electric field evaluated at the focal point at each of the three frequencies $\omega_j = \frac{2\pi c}{\lambda_j}$, which satisfies the time-harmonic Maxwell Eq. (4) in two dimensions (xy) with outgoing/radiation boundary conditions, where the current source/forcing is chosen to create a normal-incident planewave [71]. Two polarizations were considered, depending on the lens configuration, as discussed below, and we impose lengthscale constraints as in Section 3.A.

For the RGB test problem $\varepsilon(\mathbf{r})$ is interpolated linearly from the projected design field $\hat{\rho}$ (Section 3.A) between the minimum $\varepsilon_{\min} = 1$ and maximum $\varepsilon_{\max} = 5.76$ (chosen somewhat arbitrarily, but in the range of materials such as GaN and LiNbO₃). The worst-case objective Φ was reformulated into a differentiable problem by an epigraph formulation Eq. (9). The enhancement is non-dimensionalized, at each wavelength, by dividing by the intensity obtained for an empty design (that is, the intensity $|E_{\rho=0}(\mathbf{r}_0)|^2$ when $\varepsilon = 1$ everywhere in the design region). Equivalently, we normalize the incident field so that $|E_{\rho=0}(\mathbf{r}_0)|^2 = 1$.

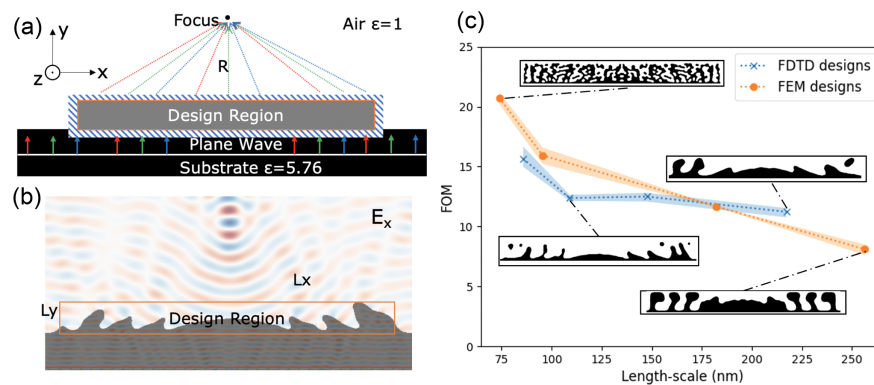


Fig. 3. RGB metalens design problem with in-plane E -field polarization (a) Schematic of the problem and setup; blue shaded region represents the padding to the design region: the bottom was padded by the substrate material and other sides were padded by air. (b) Sample design and its field pattern at $\lambda = 650$ nm. (c) Plot of FOMs versus lengthscales with insets showing some designs at various lengthscales; shading corresponds to the range of FOMs and markers correspond to the average.

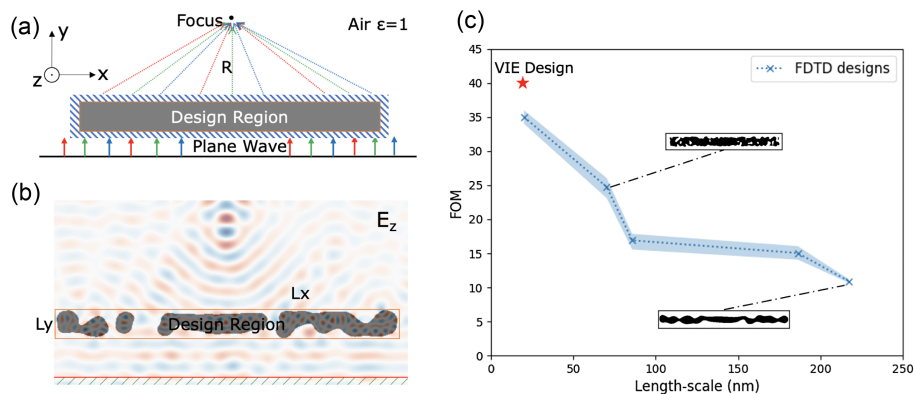


Fig. 4. RGB metalens design problem with out-of-plane E -field polarization. (a) Schematic of the problem and setup; blue shaded region represents the padding of air to all sides of the design region. (b) Sample design and its field pattern at $\lambda = 650$ nm. (c) Plot of FOMs versus lengthscales with insets showing some designs at various lengthscales; shading corresponds to the range of FOMs and markers correspond to the average; the red star indicates a VIE design without lengthscale constraint.

We considered the test problem in two different configurations: (1) in-plane E -field polarization with a substrate layer of $\epsilon = 5.76$ beneath the design [Fig. 3(a)] and (2) out-of-plane E -field polarization without a substrate beneath the design [Fig. 4(a)]. (The latter substrate-free configuration is more accessible to integral-equation methods.) We tested three independent implementations of topology optimization using three different methods for solving Maxwell's equations: (a) finite-difference time-domain (FDTD) method, implemented in the free and open-source software package Meep [16,17,51], with resolution of 50 pixels per μm , (b) frequency-domain finite-element method (FEM) implemented in MATLAB [72] with COMSOL Multiphysics [73], with resolution of 100 pixels per μm , and (c) a volume integral-equation method (VIE) [54–56], with resolution of 100 pixels per μm .

For configuration (1), we designed structures at different lengthscales separately with the FDTD and the FEM topology optimization. The performance of the optimized structures was then cross-validated (i.e., the FEM-optimized design was evaluated in FDTD and vice versa), with discrepancy found to be below 5%. An exemplary design and its electric-field pattern are shown in Fig. 3(b). We plotted FOM [figure of merit]

versus measured lengthscale in Fig. 3(c). The optimized structures display comparable performance for a given lengthscale, and gradually degrade as the minimum lengthscale increases. (Because different methods happen to find different local minima, and the solution can jump between minima as the lengthscale is changed, it is not surprising that the curves do not coincide and may not be completely smooth.)

In configuration (2), without a substrate, we again optimized structures at different lengthscales in FDTD, and cross-validated the performance with a volume-integral-equation (VIE) solver [55]. We also designed a structure using VIE, without a lengthscale constraint, and optimized for the average instead of worst-case; the resulting VIE design exhibits a measured minimum lengthscale of ≈ 20 nm. One design example and its field pattern are shown in Fig. 4(b), and FOMs versus lengthscale is plotted in Fig. 4(c). As in configuration (1), the two approaches exhibit roughly comparable performance at a given lengthscale.

In both configurations, we started the optimization with a uniform initialization at $\rho = 0.5$, but due to differences in discretization and other hyperparameters we often converged to different local optima from different simulation techniques.

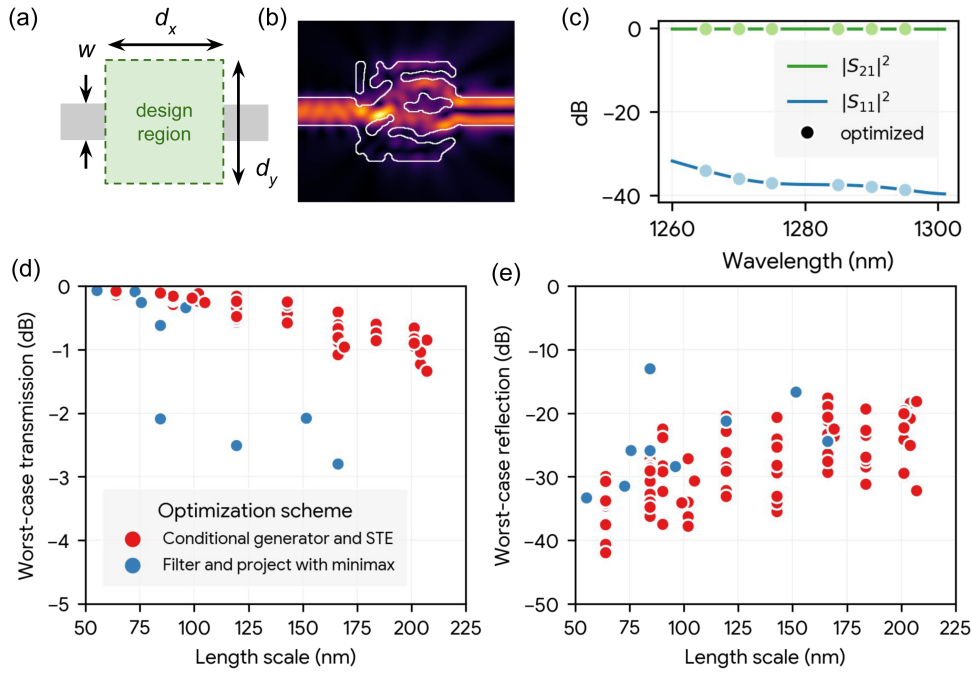


Fig. 5. (a) Schematic of a waveguide mode converter. (b) Field amplitude distribution and (c) scattering parameter spectra for an optimized mode converter. Circles indicate the wavelengths targeted during optimization. (d) Worst-case transmission/conversion efficiency and (e) worst-case reflection for a collection of mode converter designs optimized using different schemes.

Indeed, we find that this problem exhibits many local optima that achieve comparable performance; for example, we also tried $\rho = 0$ and $\rho = 1$ initializations at each lengthscale, and we observed that roughly 70% of those results were within 10%, in either direction, of those in Figs. 3(c) and 4(c), and the remaining 30% were within a factor of two worse.

C. Waveguide Mode Converter

Components for efficiently coupling light between orthogonal time-harmonic modes of optical waveguides are important for quantum information processing [74] and optical communications [75–79], where higher-order modes offer additional degrees of freedom for encoding information. Additionally, the ability to address and probe higher-order spatial modes of integrated waveguides is crucial for applications in nonlinear optics, such as sum- and difference-frequency generation [80]. Topology optimization has been applied to various mode-conversion problems in integrated optics [1,17,25–28].

As a test problem for inverse design, mode conversion probes an implementation’s ability to excite and extract different modes of a multi-mode waveguide, to characterize both transmission and reflection of individual modes, and to optimize performance over a broad continuous bandwidth (as opposed to the few discrete wavelengths of the RGB metalens test problem).

A two-dimensional test problem is depicted schematically in Fig. 5(a), showing an integrated waveguide mode-converter design problem with input and output dielectric waveguides. The design region has dimensions $1.6 \mu\text{m} \times \mu\text{m}$ with 400-nm-wide silicon ($\epsilon = 12.25$) input and output waveguides. The surrounding cladding region and the “void” material inside the design region both consist of silicon dioxide ($\epsilon = 2.25$). The electric field is polarized in the z (out-of-plane) direction. The

objective of the design problem is to convert input power from the fundamental mode of the input waveguide on the left to the second-order mode of the output waveguide on the right, with maximal efficiency (defined more precisely below) over an operating bandwidth of 1260–1300 nm. This problem is based on the waveguide mode-converter design problem presented in [52]. An optimized waveguide mode converter (with the electric-field amplitude) and its performance are shown in Figs. 5(b) and 5(c), respectively. The reflection back into the fundamental mode (S_{11}) and the transmission from the fundamental mode to the second-order mode (S_{21}) over the full operating bandwidth are shown in Fig. 5(c).

Figures 5(d) and 5(e) compare the worst-case transmission and reflection, respectively, for several designs, each optimized to satisfy varying minimum-lengthscale constraints. The objective function is the worst-case of insertion loss ($1 - |S_{21,\lambda}|^2$) and reflection ($|S_{11,\lambda}|^2$) over all sampled wavelengths (λ) within the operating bandwidth:

$$\Phi = \max_{\lambda} (1 - |S_{21,\lambda}|^2, |S_{11,\lambda}|^2). \quad (11)$$

Note that the S matrix coefficients are computed from the \mathbf{E} and \mathbf{H} fields by an overlap integral (over the port cross section) with the corresponding input/output port-mode profile \mathbf{E}_k and \mathbf{H}_k (normalized to unit power), of the form $S_k = \frac{1}{4} \int [\mathbf{E}_k^* \times \mathbf{H} + \mathbf{E} \times \mathbf{H}_k^*] \cdot d\mathbf{A}$ [81]. These are normalized by input power, so that $|S_k|^2 = 1$ corresponds to 100% transmission or reflection. Results from two different optimization and simulation algorithms are shown in Figs. 5(d) and 5(e), where each data point corresponds to the best performing feasible design from independent optimization runs. Data

marked by blue dots are from a filter-and-projection parameterization [61] scheme for density-based topology optimization (Section 3.A), with minimax optimization via the epigraph formulation of Eq. (9). These designs were optimized using the Meep FDTD (finite-difference time-domain) software [16,17]. Data marked by red dots are from a straight-through estimator (STE) and conditional-generator strategy [52] optimized using the open-source Ceviche FDFD (finite-difference frequency-domain) simulator [82,83], with a conceptually similar alternative objective function to minimize a combination of the deviation in transmission and reflection from an ideal filter [52]; the exact expression is given in Supplement 1. A 10 nm Yee grid was used for all simulations. More details on each optimization procedure are provided in Supplement 1.

Regardless of which simulator was used for optimization, the performance reported in Figs. 5(d) and 5(e) was obtained by simulating the design in Ceviche [83] and cross-validated in Meep. Across both optimization strategies, we observe a trend of gradually improving performance (higher worst-case transmission and lower worst-case reflection) for designs with smaller lengthscales, but good performance is still attained for relatively large lengthscales 100 nm. We also find that the measured lengthscale, reported on the x -axis of Figs. 5(d) and 5(e), is within 10% of the imposed lengthscale for the conditional generator + STE designs and within 20% for the filter-and-projection parameterized designs. For each lengthscale, several local optima are shown in Figs. 5(d) and 5(e) for the conditional generator + STE method. These designs have a variation of roughly 15% around the worst-case reflection averaged over all designs at each lengthscale. The variation in the worst-case transmission is 10%–30% around the average value for different lengthscales. Results are similar for local optima obtained using the filter-and-projection method via Meep FDTD.

D. Metallic Concentrator

Metallic structures designed to concentrate the electromagnetic field have many applications for the enhancement of nonlinear effects and light-matter interaction [84,85] such as Raman sensing [86,87]. As a test problem for optical design, they also exercise the ability of a solver to handle metallic materials and plasmonic resonance phenomena [88], which may require careful handling when employing density-based topology optimization for the inverse design [63]. Although particular applications are often best served by a holistic approach of optimizing the full physical process end-to-end [87], for the purpose of a test suite we consider a simplified two-dimensional problem: designing a structure that concentrates the electromagnetic field from an incident plane wave to a point in space. (Unlike the far-field focus of the RGB metalens in the previous section, this is a *near-field* focusing problem.)

Thus, in this test problem we seek to maximize the electric-field intensity at a point in space by tailoring the geometry of a metallic particle being illuminated by an electromagnetic plane wave (wavelength $\lambda = 532$ nm) propagating through free space ($\epsilon = 1$) and polarized in-plane. Like in the RGB focusing problem, we consider a 2d model problem to reduce the computational burden. (The in-plane polarization makes the problem more interesting because it allows field singularities to arise at sharp tips [89], which can have a strong impact on near-field

intensity.) To achieve point focusing, as in Section 3.B we take as the objective function

$$\Phi = |\mathbf{E}(\mathbf{r}_0)|^2, \quad (12)$$

corresponding to intensity at the focal point \mathbf{r}_0 , from an incident planewave, non-dimensionalized by dividing Φ by the intensity obtained for vacuum (no metal, i.e., the incident intensity). As we are interested in near-field focusing, we define an annulus-shaped design domain [87], centered at \mathbf{r}_0 , with outer radius $r_d = 100$ nm and inner radius of $r_i = 10$ nm, in which the metallic (silver, index $n = 0.054 + 3.429i$ [90]) particle under design is located; see Fig. 6(a). The inner radius $r_i > 0$ of the design domain is useful to regularize the problem: similar to Section 3.E, if the design were allowed to converge to an arbitrarily sharp tip at \mathbf{r}_0 , the resulting optimal Φ would diverge with increasing computational resolution. We again use density-based topology optimization, with lengthscale constraints, as in Section 3.A.

To avoid spurious resonant phenomena that can arise for metallic materials, the relative permittivity ϵ is interpolated from the projected density $\hat{\rho}$ using a formula involving the complex refractive index $n + i\kappa$, as explored in detail for topology optimization in [63]. This interpolation is given by the formula

$$\begin{aligned} \epsilon(\mathbf{r}) &= (n(\hat{\rho}(\mathbf{r}))^2 - \kappa(\hat{\rho}(\mathbf{r}))^2) - i(2n(\hat{\rho}(\mathbf{r}))\kappa(\hat{\rho}(\mathbf{r}))), \\ n(\hat{\rho}(\mathbf{r})) &= n_{\text{air}} + \hat{\rho}(\mathbf{r})(n_{\text{Ag}} - n_{\text{air}}), \\ \kappa(\hat{\rho}(\mathbf{r})) &= \kappa_{\text{air}} + \hat{\rho}(\mathbf{r})(\kappa_{\text{Ag}} - \kappa_{\text{air}}). \end{aligned} \quad (13)$$

We implement and cross-validate exemplary solutions using two independent software implementations of density-based topology optimization via the finite-element method (FEM), as described next. In both cases, the model problem is discretized using the finite-element method (FEM) and an element-wise constant ρ -field is employed. Notably, a mesh with $h = 1$ nm side-length is used to discretize the design domain: such extreme resolution is employed to accurately capture and fully exploit the near-field effects at the surface of the metal.

First, we use a COMSOL Multiphysics [73] based tool, originally developed for designing plasmonic nanoparticles to enhance single-molecule Raman scattering [87]. Here, a minimum lengthscale of 6 nm for the radius of curvature of the metallic particle is enforced. The resulting optimized metallic particle and electric-field norm for an incident plane wave at the target wavelength is shown in Fig. 6(b), clearly illustrating the focusing effect of the particle. The green circles in the panel outline the outer and inner radii of the design domain.

For cross-validation and comparison, we also used an independent implementation based on the free/open-source FEM tool Gridap [19,92], originally developed to maximize the spatially averaged (many molecule) Raman scattering [91]. In this case, no lengthscale constraint was explicitly imposed, but the inner radius $r_i > 0$ nevertheless ensures a non-diverging optimum as mentioned above. Indeed, known bounds [93,94] suggest that optimal performance may increase as $1/r_i^2$ with the inner radius r_i of the annular design region.

The two silver particles resulting from solving the design problem using the two different codes, along with a wavelength

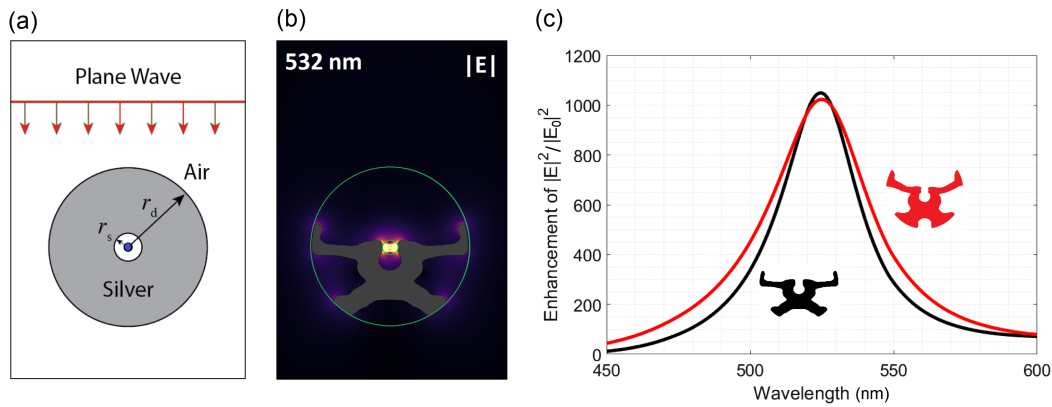


Fig. 6. Focusing of the electric field using a metallic nanoparticle. (a) Schematic of the design problem showing an incident plane wave that is to be concentrated at the center of the design region marked by the black dot. (b) $|E|$ -field using saturated colors (inferno) to illustrate the enhancement around the particle, with the particle design overlay (dark gray) at $\lambda = 532$ nm and design domain outline (green). (c) Wavelength sweep of $|E|^2$ -enhancement relative to empty space, (black) design optimized using software developed for [87], (red) design optimized using software developed for [91] freely available at [92].

sweep of the $|E(\mathbf{r}_0)|^2$ -enhancement, are shown in Fig. 6(c). Both optimized particle pixel maps were smoothed and curve fitted. The performance of the optimized designs was then validated using a high-resolution conforming mesh with second-order basis functions using COMSOL Multiphysics. Despite minor differences between the two designs, the enhancement in $|E(\mathbf{r}_0)|^2$ at the target wavelength $\lambda = 532$ nm (compared to the incident intensity) is within 5% of one another. Furthermore, we also tried starting with five different random initial structures, and all resulted in performance within 5% of each other.

E. Cavity Design

Resonances confine electromagnetic waves for a long time in a given volume of space, and are crucial to a wide variety of applications requiring enhanced light-matter interactions and/or narrow-band frequency sensitivity. A common figure of merit for a resonance is the local density of states (LDOS), corresponding to the power emitted by a point-dipole source placed at the point of interest, since the LDOS quantifies “Purcell enhancement” of spontaneous emission [71]. Several works have studied inverse design of the LDOS [34–36,95], and it poses a useful challenge as a test problem because of its *extreme* sensitivity to the minimum lengthscale. In 3d, or in 2d with the in-plane electric-field polarization, the electric field [89,96,97], and hence the LDOS, diverges at sharp tips [35,98], so optimization without a lengthscale constraint or other regularization would yield an LDOS that diverges with the computational resolution. Optimization of resonant modes is challenging for another reason: the single-frequency LDOS for a long-lifetime (or high “quality factor” Q [99]) resonance is very sensitive to small changes in geometry (which can shift the resonance off the target frequency), leading to an ill-conditioned “stiff” optimization problem that causes many algorithms to converge slowly [34]. Hence, this problem is especially useful for testing inverse-design algorithms.

In particular, we look for the dielectric profile $\varepsilon(x, y)$, controlled by the design field $\rho \in [0, 1]$ as in Section 3.A, in a

$1\lambda \times 1\lambda$ square region aiming to maximize the local density of states (LDOS) from a unit in-plane dipole source at the center (i.e., in-plane electric-field polarization) for a wavelength $\lambda = 1.55 \mu\text{m}$. A schematic is shown in Fig. 7(a). The design region’s permittivity varies from $\varepsilon_{\min} = 1.0$ to $\varepsilon_{\max} = 12.11$, and the surrounding material is $\varepsilon_{\min} = 1.0$ with outgoing boundary conditions (implemented in our case using PML). As noted above, it is crucial to impose a minimum lengthscale (or similar regularization) on the design region to avoid a diverging result, and the resulting optimized LDOS is characterized as a function of the measured minimum lengthscale (Section 2). (The finite diameter of the design region also regularizes the problem by effectively limiting the attainable Q , which would otherwise lead to a diverging LDOS for lossless materials in an unbounded design domain [34].)

The LDOS is proportional (with some normalization factor [71]) to the power radiated by the source $\sim \int \text{Re}[\mathbf{E}^* \cdot \mathbf{J}] = \text{Re}E_x(0, 0)$, where the current density $\mathbf{J} = \hat{x}\delta(x)\delta(y)$ is a delta function at the origin. ($\text{Re}E_x(0, 0)$ is finite at the origin, in contrast to $\text{Im}E_x(0, 0)$, because $\int \text{Re}[\mathbf{E}^* \cdot \mathbf{J}]$ must yield the finite radiated power in accordance with Poynting’s theorem [96].) The optimization objective is therefore simply to maximize

$$\Phi = \text{Re}E_x(0, 0),$$

and since this is strictly positive, one can equivalently minimize $1/\text{Re}E_x(0, 0)$ [34]. Our reference implementations below employed density-based topology optimization, with linear material interpolation and explicit lengthscale constraints on the solid and void regions as in Section 3.A.

We separately performed the design with FDTD in Meep [16,17,51] and FEM implemented in MATLAB [72] with COMSOL Multiphysics [73]. Simulations were performed at a resolution of 200 pixels/ μm .

The FDTD and FEM designs were cross-validated, with discrepancy below 10%. We non-dimensionalized the LDOS by dividing it by the LDOS of vacuum: although there is an analytical formula for the power emitted by a dipole in vacuum [96], it is better to calculate the vacuum LDOS numerically using the same discretization scheme with $\varepsilon = 1$, so that any

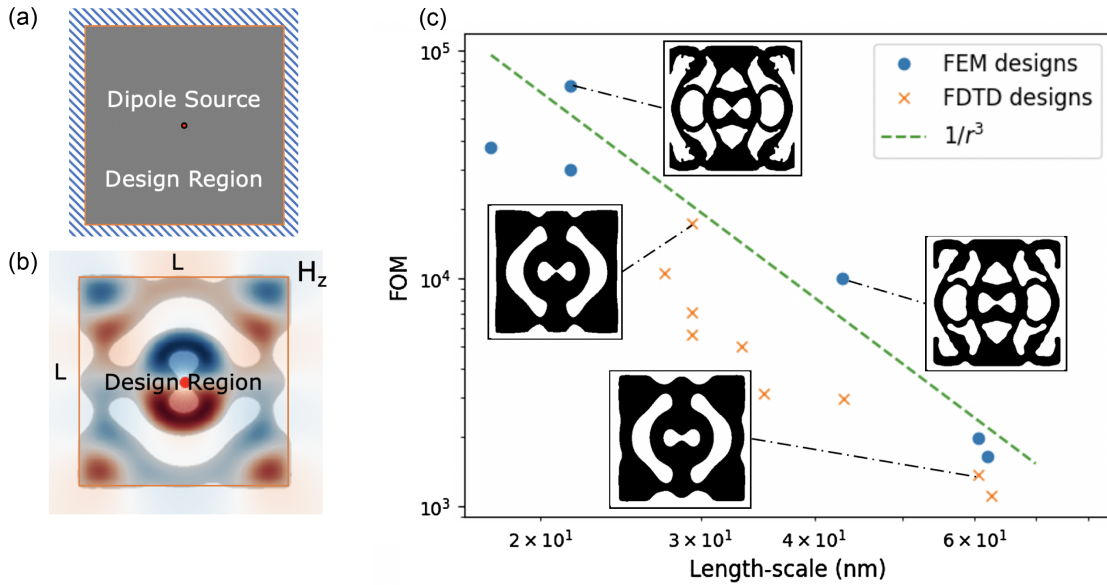


Fig. 7. Cavity design problem with E -field in-plane polarization (a) Schematic of the problem and setup; blue shaded region represents the padding of air to all sides of the design region. (b) Sample design and its H_z field pattern. (c) Plot of FOMs versus lengthscales on log-scales with insets showing some designs at various lengthscales.

normalization or units of the simulation method are cancelled. One example design and its field pattern are shown in Fig. 7(b), and FOM versus lengthscale for many designs is plotted in Fig. 7(c). We find that LDOS optimization encounters many local optima of varying performance, several of whose data points are shown in Fig. 7(c). Some of these are for different random initial ρ , but even for the same initial $\rho = 0.5$, corresponding to the inset structures, we find that the local optimum obtained is very sensitive to discretization and hyperparameters of the optimization. Nevertheless, we find that many local optima achieve a comparable LDOS, within the same order of magnitude.

While a more in-depth study of the LDOS is outside the scope of this article, it is interesting to note that Fig. 7(c) empirically exhibits an LDOS scaling roughly with $1/\text{lengthscale}^3$. Future work may relate such observations to fundamental bounds on the attainable LDOS [35,36].

F. 3d Metagrating

An important application of topology optimization is the design of a metagrating that can deflect light into a single diffraction order with high efficiency. Previous works have demonstrated freeform metagratings for large-angle, multifunctional beam deflection designed via local [29–31] and global [32,33] topology optimization. Here, we use 3d metagratings as our model system to benchmark the results between different solvers. This offers a fully 3d test problem that only requires modest computing resources, since it is periodic in the x and y directions with periods smaller than 2λ .

The metagrating consists of freeform silicon patterns (we used a resolution of 90 points per wavelength) that deflect normally incident light from a silica substrate to the +1 diffraction order in air [Fig. 8(a)]. The metasurface is periodic in both x (period $\Lambda_x = 1371$ nm) and y (period $\Lambda_y = 525$ nm)

directions, with a constant cross section in z (consisting of a 325-nm-thick silicon layer on top of a semi-infinite silica substrate, with $\epsilon_{\text{silicon}} = 3.45^2$ and $\epsilon_{\text{silica}} = 1.45^2$, respectively). The figure of merit of this problem is the diffraction efficiency to the $(m_x, m_y) = (+1, 0)$ diffraction channel, corresponding to a 50° deflection angle in the xz plane for the chosen wavelength and period. Here, m_x and m_y denote the diffraction orders that scatter the in-plane wavevector components (k_x, k_y) from $(0,0)$ to $(k_x + m_x 2\pi/\Lambda_x, k_y + m_y 2\pi/\Lambda_y)$. The diffraction efficiency is defined as the intensity (power/area) of light deflected to the desired diffraction order, normalized to the light intensity incident from within the silica substrate, for the p polarization (incident electric field in the x direction) at a vacuum wavelength $\lambda = 1050$ nm. The optimization objective can therefore be formulated as

$$\Phi = |S_{(+1,0)}|^2, \quad (14)$$

where $S_{(+1,0)}$ is the amplitude of the output diffraction channel (whose squared magnitude is proportional to power per unit cell in that channel), computed by an overlap integral with \mathbf{E} and \mathbf{H} as in Section 3.C and normalized so that $\Phi = 1$ corresponds to 100% conversion of the incident power.

We applied two independent density-based topology-optimization implementations to this problem, as in Section 3.A. First, we employed a free/open-source rigorous coupled-wave analysis (RCWA) solver [57] (also called a modal-expansion method) with a resolution of roughly 90 points per wavelength (i.e., a pixel size of roughly 12 nm) in the xy plane. Second, we employed the Meep FDTD simulation and topology-optimization method [16,17,51], both for cross-validating the RCWA results and for an independently optimized design. In general, this problem exhibits a large number of distinct local optima that yield similar efficiencies ($>90\%$).

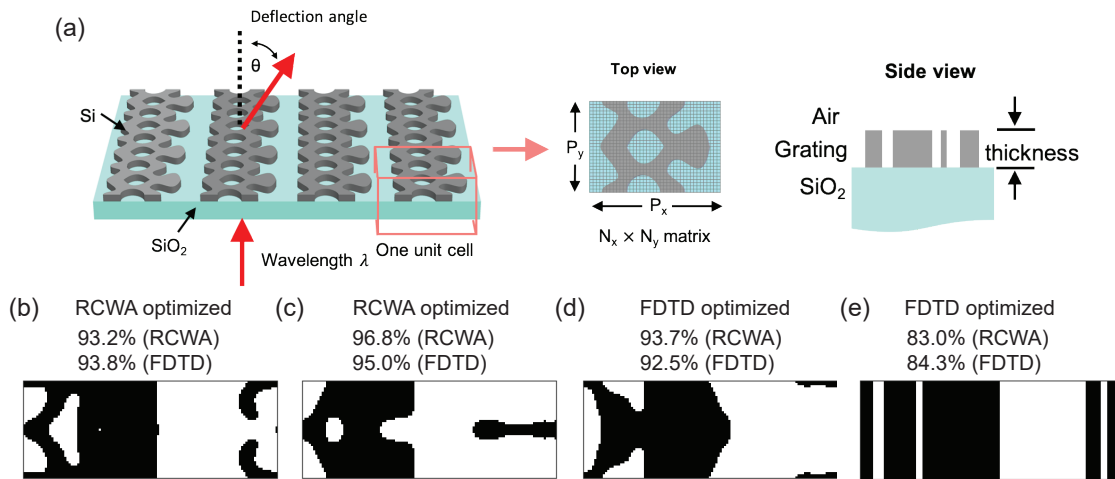


Fig. 8. 3d metagrating deflector designs. (a) Schematic of the problem setup. The metagratings consist of freeform silicon patterns and deflect normally incident light to the +1 diffraction order. (b)–(e) Patterns of refractive indices for four structures optimized by Reticolo RCWA (b) and (c) and Meep FDTD (d) and (e). The black and white regions represent silicon and air, respectively.

To reduce device sensitivity to fabrication imperfections, the RCWA approach employed a form of “robust” optimization [100]: at each iteration, the optimization maximizes the average performance of nominal, dilated, and eroded designs. As in Section 3.A, we filter and project the design field ρ , but now the projection of Eq. (7) is done in three ways: projecting with $\eta = 0.5$ yields the ideal (nominal) design pattern $\hat{\rho}_i$. Eroded and dilated patterns, denoted by $\hat{\rho}_e$ and $\hat{\rho}_d$, are generated from the same projection with η values above and below 0.5, respectively. The filter radius r_f and the η parameters are related to the extent of erosion and dilation, which is described by the edge deviation [100]. In this work, the edge deviation is set as 5 nm. Given the diffraction efficiencies η of the eroded, dilated, and nominal patterns, we optimized the sum of $\eta(1 - \eta/2)$ for the three designs, which corresponds to averaging the gradients of η with weights $1 - \eta$ (so that the evolution of the design is dominated by the lowest- η variant). The free/open-source code and more details of the RCWA-based algorithm are available at [101,102].

The metagrating problem (whether or not robustness is included) exhibits many distinct local optima, but they typically have similar diffraction efficiencies, which are $>90\%$ for 2d design patterns and $\sim 80\%$ for 1d design patterns. Starting from randomly initialized geometries, the permittivity profiles of two optimized metagratings in a unit cell are shown in Figs. 8(b) and 8(c). The two metagratings have diffraction efficiencies of 93.2% and 96.8%; this performance is also cross-validated by Meep, which reports 93.8% and 95.0%, respectively, for the same structures. We also performed topology optimization using Meep [17] to maximize the nominal efficiency η (no robustness, i.e., no eroded or dilated structures). The Meep-optimized structures, evolved from a homogeneous starting design with $\varepsilon = (\varepsilon_{\text{silicon}} + 1)/2$, are shown in Figs. 8(d) and 8(e). For the design in Fig. 8(d), the diffraction efficiencies at $\lambda = 1050$ nm evaluated by FDTD and RCWA are 92.5% and 93.7%, respectively. To illustrate the effect of reducing the number of degrees of freedom, we also include a design in Fig. 8(e) in which the density $\rho(x)$ was restricted to vary only along the x direction, resulting in a 1d grating structure with a

diffraction efficiency at $\lambda = 1050$ nm of 84.3% (cross-validated in RCWA as 83.0%) starting from a homogeneous structure with $\varepsilon = (\varepsilon_{\text{silicon}} + 1)/2$. All these designs were optimized without lengthscale constraints. The estimated minimum lengthscales of the designs in Figs. 8(b)–8(e) are 66, 62, 54, and 35 nm, respectively. [As explained in Section 2, our lengthscale metric disregards features at the 1–2 pixel level as discretization artifacts; for example, the single white pixel visible in Fig. 8(b) is ignored.] Similar to Section 3.C (since a diffraction grating is mathematically a form of mode converter), however, we expect the performance to degrade only gradually as moderate lengthscale constraints are imposed.

G. Dispersion Engineering

Topology-optimization methods have been widely employed to tailor the dispersion relation of the photonic crystals, including enlarging bandgaps [2,103], reducing the group-velocity dispersion (GVD) of slow-light waveguides [37,38], and obtaining nontrivial topological properties [104,105]. Mathematically, such tasks impose unique challenges because they involve optimizing the solution of eigenvalue problems, as opposed to linear scattering equations as in the previous sections. Therefore, we include an exemplary dispersion engineering test problem involving the design of broadband slow-light (low-group-velocity) waveguides.

In particular, we consider waveguide modes confined by the photonic bandgap of a photonic crystal (PhC) consisting of a periodic array (triangular lattice) of air ($\varepsilon = 1$) holes in silicon ($\varepsilon = 12.08$) for the H_z (“TE”) polarization in 2d [99]. We optimize a design region forming a “line defect” that breaks the periodicity along one direction (orthogonal to the waveguide, denoted y) and thereby introduces one or more localized modes into the bandgap [99]. These guided modes are found via the band structure calculations of the periodic $a \times b$ supercell [see Fig. 9(a), Si = black, air = white]. The H_z eigenfunctions in the supercell solve Eq. (5) as an eigenvalue problem by taking $\mathbf{K}(\mathbf{r}) = 0$ [99], subject to Bloch-Floquet boundary conditions

with a Bloch wavevector $(k, 0)$ stated as

$$H_z(x, a) = e^{ika} H_z(x, 0), \quad H_z(0, y) = H_z(b, y). \quad (15)$$

For a given eigenvalue ω , the dispersion $\omega(k)$ has a propagation velocity given by the group velocity v_g [99]:

$$v_g = \frac{c}{n_g} = -\frac{\partial \omega}{\partial k}, \quad (16)$$

where n_g defines the “group index” of the guided mode.

Here, our slow-light design problem is to obtain a nearly constant group index n_g matching a target $n_g^* = 25$ over a given bandwidth $\Delta\omega$, or equivalently over a given range of wavenumbers $\Delta k = \Delta\omega n_g^*/c$. More precisely, we minimize the maximum squared error $(n_g - n_g^*)^2$ over a given set of Bloch wavenumbers k_i , where for convenience we approximate n_g by a finite-difference approximation for adjacent k_i 's (one can alternatively use analytical differentiation with k [106]). We also include band-separation constraints to keep the targeted modes gap-guided as in [38]. This results in the optimization problem

$$\begin{aligned} \min_{\rho} \quad & \max_{\eta, i \in \{2, \dots, m\}} \left(\frac{c(k_i - k_{i-1})}{\omega_n^\eta(k_{i-1}) - \omega_n^\eta(k_i)} - n_g^* \right)^2 \\ \text{s.t.} \quad & \max_j \omega_{n-1}^\eta(k_j) \leq a_1 \min_i \omega_n^\eta(k_i), \\ & \omega_n^\eta(0) \geq a_2 \max_i \omega_n^\eta(k_i), \\ & \min_j \omega_{n+1}^\eta(k_j) \geq a_2 \max_i \omega_n^\eta(k_i), \end{aligned} \quad (17)$$

where $\rho(x, y) \in [0, 1]$ is the design density (discretized as described below), $\eta \in \{\eta_d, \eta_i, \eta_e\}$ represents three design realizations with different thresholds (see below) to improve the design robustness to geometric imperfections with η_d , η_i , and η_e corresponding to the dilated, normal, and eroded designs, k_i denotes the target wavenumber at each point i , while $a_1 < 1$ and $a_2 > 1$ are the band-separation factors (forcing ω_n to lie above ω_{n-1} by a factor $1/a_1$ and below ω_{n+1} by a factor $1/a_2$).

In our exemplary implementation, the finite-element method is employed to solve the eigenvalue problem. The supercell in Fig. 9(a) is discretized with 40×408 square four-node elements (40 elements/ a). Figures 9(c), 9(e), and 9(f) show the band structure and corresponding group index for the even-symmetry guided modes, along with the H_z pattern of a guided mode. As the edge $k = \pi/a$ of the Brillouin zone is approached, the group velocity goes to zero ($n_g \rightarrow \infty$), and the group-velocity dispersion diverges—this makes it challenging to operate slow-light devices without signal distortion [107], and makes it desirable to optimize a structure for nearly constant n_g (low dispersion) over a wide k band.

A linear interpolation scheme for the inverse of relative permittivity is employed, generated using a threshold value $\eta \in \{\eta_d, \eta_i, \eta_e\}$, given as

$$\frac{1}{\varepsilon^\eta(x, y)} = \frac{1}{\varepsilon_{\text{air}}} + \hat{\rho}^\eta(x, y) \left(\frac{1}{\varepsilon_{\text{Si}}} - \frac{1}{\varepsilon_{\text{air}}} \right). \quad (18)$$

Here, a supercell is designed for a target group index of $n_g^* = 25$ with seven equidistant points in the target wavenumber

interval $k \in [0.3875, 0.4625] \times 2\pi/a$ and band-separation factors of $a_1 = 0.9$ and $a_2 = 1.1$. Similar to Section 3.F, three design realizations are generated using projection thresholds $\eta \in [0.35, 0.5, 0.65]$ with a filter radius $r_f = a/8$. The design domain is illustrated in Fig. 9(b). Figures 9(d), 9(e), and 9(g) show the band structure, group index, and a guided mode for the nominal optimized design ($\eta = 0.5$). Compared to the initial design, the optimized design possesses a near-constant group index profile in the target wavenumber interval.

A smoothed design is extracted using a contour to generate a high-resolution design with 160 elements/ a . The high-resolution design was cross-checked with commercial software COMSOL [73] and with the free/open-source MPB [18], and we found that the low-resolution optimized design [insert in Fig. 9(d)] was accurate to within 0.3% in frequency and within 3.5% in the group index.

Although we did not explicitly constrain the lengthscale in this optimized design, a minimum lengthscale ($0.0968a$) is effectively imposed via the three-case robust optimization, and the final result has a minimum lengthscale of $0.113a$ according to the metric in Section 2.

In this study, it is important to choose a proper initial design so that the band-separation constraints can be satisfied in the optimization procedure, i.e., so that the starting guess is feasible, since otherwise the optimization algorithm is not guaranteed to converge [66]. One cannot simply choose a random initial ρ . Initial designs with different r_2 ranging from $0.2a$ to $0.275a$ have been tested. They all result in similar designs with maximum objective value deviations within 1.5%.

4. CONCLUDING REMARKS

Although the test problems and exemplary solutions in the preceding structures are mostly based on results from a variety of publications in the literature, this paper represents more than a mere compilation. By systematically cross-checking multiple software packages and optimization algorithms for given structures and problems in topology optimization, characterizing results via an independent lengthscale measure, and providing reproducible data and scripts (in Dataset 1, Ref. [13]), we hope that this work represents a foundation that can add confidence and speed development of future progress in large-scale optimization for photonics.

An important complement to a validation suite is a benchmark suite, in order to compare the computational efficiencies of different inverse-design approaches, but we have intentionally avoided the latter topic in the present work. Benchmarking is subtle and complicated, and great care must be taken in determining precisely what metric is being compared [12]. Indeed, it is well known that two implementations of the “same” mathematical algorithm may have runtimes that differ by an order of magnitude or more due merely to software-implementation issues, so often it is desirable to compare efficiency using some other measure such as the number of forward/adjoint Maxwell solvers to reach a given objective value, but that raises its own difficulties when disparate solvers (e.g., time- versus frequency-domain) are to be compared. Moreover, the greatest demands on computational efficiency arise when inverse-design is pushed

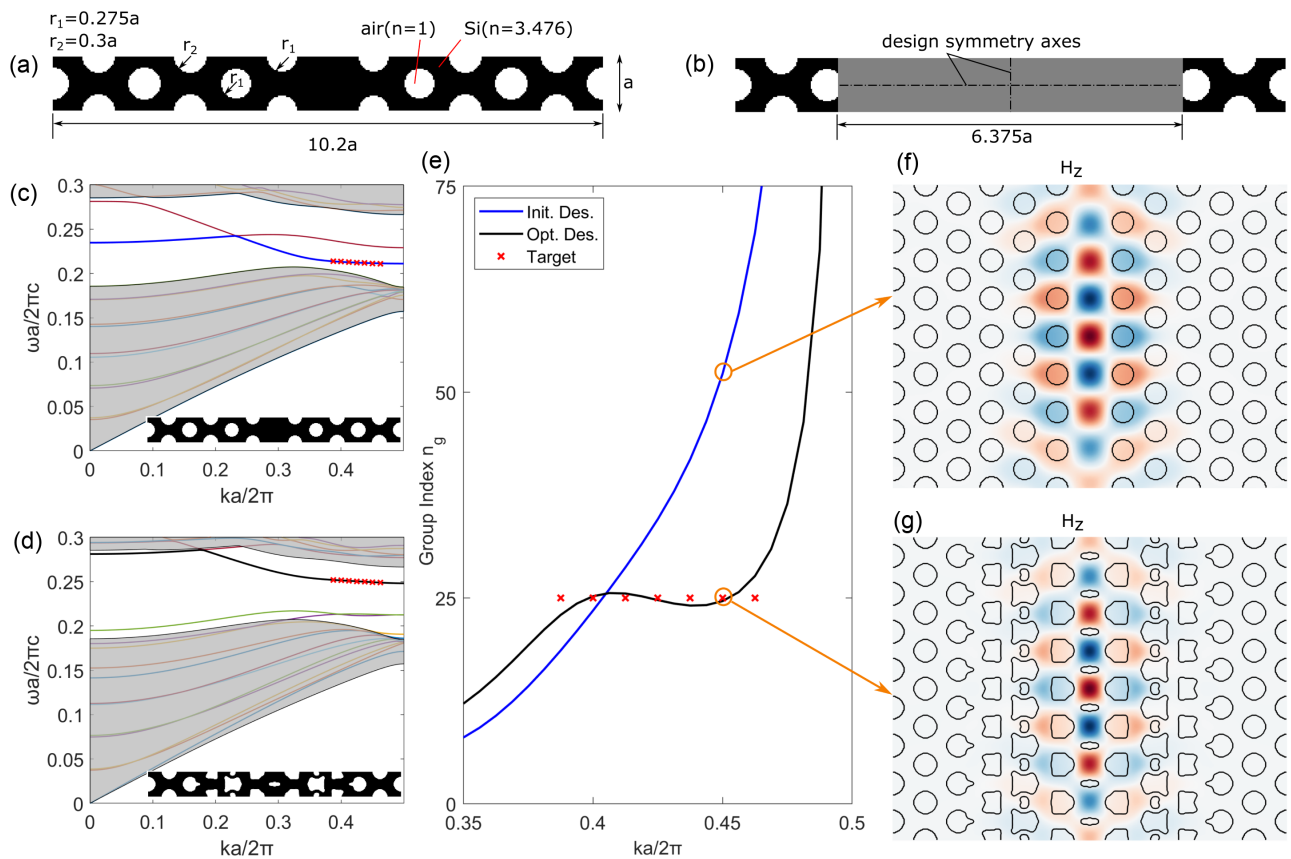


Fig. 9. Dispersion engineering for slow-light waveguides. Top: (a) initial design and (b) design domain. Bottom: band structures of the (c) initial and (d) optimized designs for a target group index of $n_g^* = 25$; group index n_g versus k ; H_z field in the initial and optimized designs.

forward towards more complicated physical settings, more difficult objectives, and ever larger scales. So, the very accessibility and common-denominator nature of validation suites may make them less interesting for exploring the limits of computational efficiency. Nevertheless, some aspects of the present work may be useful for future efforts on performance benchmarking, since of course any benchmarked algorithm must first be validated, and benchmarked algorithms must attain a common lengthscale metric to be comparable.

Funding. X, The Moonshot Factory; Simons Foundation; Danmarks Grundforskningsfond (DNRF147); Thomas B. Thrighes Foundation; National Science Foundation (2103301); Army Research Office (W911NF-18-2-0048).

Acknowledgment. All authors contributed to the overall ideas and writing of this paper. Due to the large collaboration involved in this work, we listed authors alphabetically above and note some of the specific author contributions here. REC designed one metallic concentrator and a set of RGB metalenses of Sections 3.D and 3.B, respectively, using a FEM-based TopOpt package. GI designed one set of cavities of Section 3.E using TopOpt. YZ, JJ, and JF designed the 3d metagratings of Section 3.F using the RCWA solver Reticolo. AO and WM designed the 3d metagratings of Section 3.F using the FDTD solver Meep. MC designed a set of RGB metalenses and LDOS cavities of Sections 3.B and 3.E, respectively, using Meep. WX and ODM designed the RGB metalens of Section 3.B using a custom VIE solver. FW designed the slow-light waveguides of Section 3.G using TopOpt. SGJ provided overall guidance and validated the slow-light waveguide designs of Section 3.G using the frequency-domain plane-wave-expansion solver MPB. IADW and MFS designed the waveguide mode converters of Section 3.C using the FDFD solver Ceviche. AO designed the waveguide mode converter of Section 3.C using

Meep. WM, IADW, MFS, AO, and SGJ contributed to the development of the algorithm and software for the `Imageruler` package of Section 2.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper can be found in part in [Dataset 1](#), Ref. [13], for the optimized designs and corresponding values of objective functions; other data are not publicly available at this time but may be obtained from the authors upon reasonable request.

Supplemental document. See [Supplement 1](#) for supporting content.













REFERENCES

1. S. Molesky, Z. Lin, A. Y. Piggott, *et al.*, "Inverse design in nanophotonics," *Nat. Photonics* **12**, 659–670 (2018).
2. J. S. Jensen and O. Sigmund, "Topology optimization for nano-photonics," *Laser Photonics Rev.* **5**, 308–321 (2011).
3. G. Angeris, J. Vučković, and S. Boyd, "Heuristic methods and performance bounds for photonic design," *Opt. Express* **29**, 2827–2854 (2021).
4. P. Chao, B. Streckha, R. K. Defo, *et al.*, "Physical limits in electromagnetism," *Nat. Rev. Phys.* **4**, 543–559 (2022).
5. O. D. Miller, "Fundamental limits to near-field optical response," in *Advances in Near-Field Optics*, R. Gordon, ed. (Springer, 2023), pp. 25–85.
6. M. Zhou, B. S. Lazarov, F. Wang, *et al.*, "Minimum length scale in topology optimization by geometric constraints," *Comput. Method. Appl. Mech. Eng.* **293**, 266–282 (2015).
7. B. S. Lazarov, F. Wang, and O. Sigmund, "Length scale and manufacturability in density-based topology optimization," *Arch. Appl. Mech.* **86**, 189–218 (2016).

8. L. Hägg and E. Wadbro, "On minimum length scale control in density based topology optimization," *Struct. Multidiscip. Optim.* **58**, 1015–1032 (2018).
9. J. V. Carstensen and J. K. Guest, "Projection-based two-phase minimum and maximum length scale control in topology optimization," *Struct. Multidiscip. Optim.* **58**, 1845–1860 (2018).
10. D. Vercautse, N. V. Sapra, L. Su, *et al.*, "Analytical level set fabrication constraints for inverse design," *Sci. Rep.* **9**, 8999 (2019).
11. S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University, 2004).
12. O. Sigmund, "On benchmarking and good scientific practise in topology optimization," *Struct. Multidiscip. Optim.* **65**, 315 (2022).
13. M. Chen, R. E. Christiansen, J. A. Fan, *et al.*, "Dataset1," figshare (2023), <https://doi.org/10.6084/m9.figshare.25043612>.
14. A. Oskooi, "Imageruler," GitHub (2023), <https://github.com/NanoComp/imageruler>.
15. S. G. Johnson, "Photonics-opt-testbed," GitHub (2023), <https://github.com/NanoComp/photonics-opt-testbed>.
16. A. F. Oskooi, D. Roundy, M. Ibanescu, *et al.*, "Meep: a flexible free-software package for electromagnetic simulations by the FDTD method," *Comput. Phys. Commun.* **181**, 687–702 (2010).
17. A. M. Hammond, A. Oskooi, M. Chen, *et al.*, "High-performance hybrid time/frequency-domain topology optimization for large-scale photonics inverse design," *Opt. Express* **30**, 4467–4491 (2022).
18. S. G. Johnson and J. D. Joannopoulos, "Block-iterative frequency-domain methods for Maxwell's equations in a planewave basis," *Opt. Express* **8**, 173–190 (2001).
19. S. Badia and F. Verdugo, "Gridap: an extensible finite element toolbox in Julia," *J. Open Source Softw.* **5**, 2520 (2020).
20. A. R. Kan and G. T. Timmer, "Stochastic global optimization methods part i: clustering methods," *Math. Program.* **39**, 27–56 (1987).
21. P. Azunre, J. Jean, C. Rotschild, *et al.*, "Guaranteed global optimization of thin-film optical systems," *New J. Phys.* **21**, 073050 (2019).
22. P. S. Carney and J. C. Schotland, "Inverse scattering for near-field microscopy," *Appl. Phys. Lett.* **77**, 2798–2800 (2000).
23. G. Bao and P. Li, "Inverse medium scattering problems for electromagnetic waves," *SIAM J. Appl. Math.* **65**, 2049–2066 (2005).
24. Z. Li, R. Pestourie, Z. Lin, *et al.*, "Empowering metasurfaces with inverse design: principles and applications," *ACS Photonics* **9**, 2178–2192 (2022).
25. J. Lu and J. Vučković, "Objective-first design of high-efficiency, small-footprint couplers between arbitrary nanophotonic waveguide modes," *Opt. Express* **20**, 7221–7236 (2012).
26. L. F. Frellsen, Y. Ding, O. Sigmund, *et al.*, "Topology optimized mode multiplexing in silicon-on-insulator photonic wire waveguides," *Opt. Express* **24**, 16866–16873 (2016).
27. F. Callewaert, S. Butun, Z. Li, *et al.*, "Inverse design of an ultra-compact broadband optical diode based on asymmetric spatial mode conversion," *Sci. Rep.* **6**, 32577 (2016).
28. C. Shang, J. Yang, A. M. Hammond, *et al.*, "Inverse-designed lithium niobate nanophotonics," *ACS Photonics* **10**, 1019–1026 (2023).
29. D. Sell, J. Yang, S. Doshay, *et al.*, "Large-angle, multifunctional metagratings based on freeform multimode geometries," *Nano Lett.* **17**, 3752–3757 (2017).
30. D. Sell, J. Yang, E. W. Wang, *et al.*, "Ultra-high-efficiency anomalous refraction with dielectric metasurfaces," *ACS Photonics* **5**, 2402–2407 (2018).
31. D. Sell, J. Yang, S. Doshay, *et al.*, "Periodic dielectric metasurfaces with high-efficiency, multiwavelength functionalities," *Adv. Opt. Mater.* **5**, 1700645 (2017).
32. J. Jiang and J. A. Fan, "Simulator-based training of generative neural networks for the inverse design of metasurfaces," *Nanophotonics* **9**, 1059–1069 (2020).
33. J. Jiang and J. A. Fan, "Global optimization of dielectric metasurfaces using a physics-driven neural network," *Nano Lett.* **19**, 5366–5372 (2019).
34. X. Liang and S. G. Johnson, "Formulation for scalable optimization of microcavities via the frequency-averaged local density of states," *Opt. Express* **21**, 30812–30841 (2013).
35. O. D. Miller, A. G. Polimeridis, M. T. H. Reid, *et al.*, "Fundamental limits to optical response in absorptive systems," *Opt. Express* **24**, 3329–3364 (2016).
36. G. Işıklar, P. T. Kristensen, J. Mørk, *et al.*, "On the trade-off between mode volume and quality factor in dielectric nanocavities optimized for purcell enhancement," *Opt. Express* **30**, 47304–47314 (2022).
37. R. Stainko and O. Sigmund, "Tailoring dispersion properties of photonic crystal waveguides by topology optimization," *Waves Random Complex Medium* **17**, 477–489 (2007).
38. F. Wang, J. S. Jensen, and O. Sigmund, "Robust topology optimization of photonic crystal waveguides with tailored dispersion properties," *J. Opt. Soc. Am. B* **28**, 387–397 (2011).
39. M. Köfferlein, "Klayout," Klayout, 2020, <https://www.klayout.de/>.
40. Gdsfactory, "Gdsfactory 7.10.5," GitHub 2022, <https://gdsfactory.github.io/gdsfactory/>.
41. Y. Elesin, B. Lazarov, J. Jensen, *et al.*, "Time domain topology optimization of 3D nanophotonic devices," *Photonics Nanostruct. Fundam. Appl.* **12**, 23–33 (2014).
42. Z. Lin, C. Roques-Carnes, R. Pestourie, *et al.*, "End-to-end nanophotonic inverse design for imaging and polarimetry," *Nanophotonics* **10**, 1177–1187 (2020).
43. W. Yao, M. Benzaouia, O. D. Miller, *et al.*, "Approaching the upper limits of the local density of states via optimized metallic cavities," *Opt. Express* **28**, 24185–24197 (2020).
44. C. Roques-Carnes, Z. Lin, R. E. Christiansen, *et al.*, "Toward 3d-printed inverse-designed metaoptics," *ACS Photonics* **9**, 43–51 (2022).
45. J. Gedeon, E. Hassan, and A. C. Lesina, "Time-domain topology optimization of arbitrary dispersive materials for broadband 3d nanophotonics inverse design," *ACS Photonics* **10**, 3875–3887 (2023).
46. G. Roberts, C. Ballew, T. Zheng, *et al.*, "3d-patterned inverse-designed mid-infrared metaoptics," *Nat. Commun.* **14**, 2768 (2023).
47. R. Gonzalez and R. Woods, *Digital Image Processing*, 4th ed. (Pearson, 2017), chap. 9.
48. G. X. Ritter and J. N. Wilson, *Handbook of Computer Vision Algorithms in Image Algebra* (CRC Press, 2000), chap. 7.
49. O. Sigmund, "Morphology-based black and white filters for topology optimization," *Struct. Multidiscip. Optim.* **33**, 401–424 (2007).
50. P. Soille, *Morphological Image Analysis Principles and Applications*, 2nd ed. (Springer-Verlag, 2004), chaps. 2, 3.
51. A. M. Hammond, A. Oskooi, S. G. Johnson, *et al.*, "Photonic topology optimization with semiconductor-foundry design-rule constraints," *Opt. Express* **29**, 23916–23938 (2021).
52. M. F. Schubert, A. K. C. Cheung, I. A. D. Williamson, *et al.*, "Inverse design of photonic devices with strict foundry fabrication constraints," *ACS Photonics* **9**, 2327–2336 (2022).
53. G. Bradski, "The OpenCV Library," Dr. Dobb's Journal of Software Tools (2000).
54. J.-M. Jin, *Theory and Computation of Electromagnetic Fields* (Wiley, 2015).
55. W. Xue, H. Zhang, A. Gopal, *et al.*, "Fullwave design of cm-scale cylindrical metasurfaces via fast direct solvers," *arXiv*, arXiv:2308.08569 (2023).
56. P.-G. Martinsson, *Fast Direct Solvers for Elliptic PDEs* (Society for Industrial and Applied Mathematics, 2019).
57. J. P. Hugonin and P. Lalanne, "Reticolo software for grating analysis," *arXiv*, arXiv:2101.00901 (2021).
58. J.-P. Berenger, "A perfectly matched layer for the absorption of electromagnetic waves," *J. Comput. Phys.* **114**, 185–200 (1994).
59. Z. Sacks, D. Kingsland, R. Lee, *et al.*, "A perfectly matched anisotropic absorber for use as an absorbing boundary condition," *IEEE Trans. Antennas Propag.* **43**, 1460–1463 (1995).
60. A. Oskooi and S. G. Johnson, "Distinguishing correct from incorrect PML proposals and a corrected unsplit PML for anisotropic, dispersive media," *J. Comput. Phys.* **230**, 2369–2377 (2011).
61. F. Wang, B. S. Lazarov, and O. Sigmund, "On projection methods, convergence and robust formulations in topology optimization," *Struct. Multidiscip. Optim.* **43**, 767–784 (2011).
62. B. S. Lazarov and O. Sigmund, "Filters in topology optimization based on Helmholtz-type differential equations," *Int. J. Numer. Methods Eng.* **86**, 765–781 (2010).

63. R. E. Christiansen, J. Vester-Petersen, S. P. Madsen, *et al.*, "A non-linear material interpolation for design of metallic nano-particles using topology optimization," *Comput. Method. Appl. M.* **343**, 23–39 (2019).
64. R. E. Christiansen and O. Sigmund, "Inverse design in photonics by topology optimization: tutorial," *J. Opt. Soc. Am. B* **38**, 496–509 (2021).
65. O. D. Miller, "Photonic design: from fundamental solar cell physics to computational inverse design," *arXiv*, arXiv:1308.0212 (2013).
66. K. Svanberg, "A class of globally convergent optimization methods based on conservative convex separable approximations," *SIAM J. Optim.* **12**, 555–573 (2002).
67. M. Khorasaninejad, W. T. Chen, R. C. Devlin, *et al.*, "Metalenses at visible wavelengths: diffraction-limited focusing and subwavelength resolution imaging," *Science* **352**, 1190–1194 (2016).
68. Z. Lin, B. Groever, F. Capasso, *et al.*, "Topology-optimized multilayered metaoptics," *Phys. Rev. Appl.* **9**, 044030 (2018).
69. R. E. Christiansen, Z. Lin, C. Roques-Carnes, *et al.*, "Fullwave Maxwell inverse design of axisymmetric, tunable, and multi-scale multi-wavelength metalenses," *Opt. Express* **28**, 33854–33868 (2020).
70. H. Chung and O. D. Miller, "High-NA achromatic metalenses by inverse design," *Opt. Express* **28**, 6945–6965 (2020).
71. A. Oskooi and S. G. Johnson, "Electromagnetic wave source conditions," in *Advances in FDTD Computational Electrodynamics: photonics and Nanotechnology*, A. Taflov, A. Oskooi, and S. G. Johnson, eds. (Artech, 2013), chap. 4, pp. 65–100.
72. MathWorks, "Matlab version 9.9.0 (r2020b)," (2020).
73. COMSOL AB, "COMSOL multiphysics reference manual," Stockholm, Sweden, 2020.
74. A. Mohanty, M. Zhang, A. Dutt, *et al.*, "Quantum interference between transverse spatial waveguide modes," *Nat. Commun.* **8**, 14010 (2017).
75. B. Stern, X. Zhu, C. P. Chen, *et al.*, "On-chip mode-division multiplexing switch," *Optica* **2**, 530–535 (2015).
76. K. Y. Yang, C. Shirpurkar, A. D. White, *et al.*, "Multi-dimensional data transmission using inverse-designed silicon photonics and microcombs," *Nat. Commun.* **13**, 7862 (2022).
77. V. Liu, D. A. B. Miller, and S. Fan, "Ultra-compact photonic crystal waveguide spatial mode converter and its connection to the optical diode effect," *Opt. Express* **20**, 28388–28397 (2012).
78. A. Melikyan and P. Dong, "Adiabatic mode converters for silicon photonics: power and polarization broadband manipulators," *APL Photonics* **4**, 030803 (2019).
79. Z. Tao, B. Wang, B. Bai, *et al.*, "An ultra-compact polarization-insensitive slot-strip mode converter," *Front. Optoelectron.* **15**, 5 (2022).
80. R. W. Boyd, *Nonlinear Optics* (Elsevier, 2020).
81. D. Marcuse, *Theory of Dielectric Optical Waveguides*, 2nd ed. (Academic, 1991).
82. T. W. Hughes, I. A. D. Williamson, M. Minkov, *et al.*, "Forward-mode differentiation of Maxwell's equations," *ACS Photonics* **6**, 3010–3016 (2019).
83. I. Williamson, "Ceviche challenges: photonic inverse design suite," GitHub (2022), <https://github.com/google/ceviche-challenges>.
84. J. A. Schuller, E. S. Barnard, W. Cai, *et al.*, "Plasmonics for extreme light concentration and manipulation," *Nat. Mater.* **9**, 193–204 (2010).
85. S. P. Madsen, J. Christiansen, R. E. Christiansen, *et al.*, "Improving the efficiency of upconversion by light concentration using nanoparticle design," *J. Phys. D* **53**, 073001 (2020).
86. D. A. Long, *Raman Spectroscopy* (McGraw Hill Higher Education, 1977).
87. R. E. Christiansen, J. Michon, M. Benzaouia, *et al.*, "Inverse design of nanoparticles for enhanced Raman scattering," *Opt. Express* **28**, 4444–4462 (2020).
88. H. Raether, *Surface Plasmons on Smooth and Rough Surfaces and on Gratings* (Springer-Verlag, 1988).
89. J. Andersen and V. Solodukhov, "Field behavior near a dielectric wedge," *IEEE Trans. Antennas Propag.* **26**, 598–602 (1978).
90. P. B. Johnson and R. W. Christy, "Optical constants of noble metals," *Phys. Rev. B* **6**, 4370–4379 (1972).
91. W. Yao, F. Verdugo, H. O. Everitt, *et al.*, "Designing structures that maximize spatially averaged surface-enhanced Raman spectra," *Opt. Express* **31**, 4964–4977 (2023).
92. W. Yao, "Source code for spatially averaged SERS enhancement," GitHub (2022), <https://github.com/WenjieYao/RamanInFluid>.
93. H. Shim, L. Fan, S. G. Johnson, *et al.*, "Fundamental limits to near-field optical response over any bandwidth," *Phys. Rev. X* **9**, 011043 (2019).
94. P. Chao, R. Kuate Defo, S. Molesky, *et al.*, "Maximum electromagnetic local density of states via material structuring," *Nanophotonics* **12**, 549–557 (2022).
95. F. Wang, R. E. Christiansen, Y. Yu, *et al.*, "Maximizing the quality factor to mode volume ratio for ultra-small photonic crystal cavities," *Appl. Phys. Lett.* **113**, 241101 (2018).
96. J. D. Jackson, *Classical Electrodynamics*, 3rd ed. (Wiley, 1998).
97. V. V. Klimov, M. Ducloy, and V. S. Letokhov, "Spontaneous emission of an atom in the presence of nanobodies," *Quantum Electron.* **31**, 569–586 (2001).
98. H. Choi, M. Heuck, and D. Englund, "Self-similar nanocavity design with ultrasmall mode volume for single-photon nonlinearities," *Phys. Rev. Lett.* **118**, 223605 (2017).
99. J. Joannopoulos, S. Johnson, J. Winn, *et al.*, *Photonic Crystals: molding the Flow of Light*, 2nd ed. (Princeton University, 2011).
100. E. W. Wang, D. Sell, T. Phan, *et al.*, "Robust design of topology-optimized metasurfaces," *Opt. Mater. Express* **9**, 469–482 (2019).
101. Jonathan Fan Group, "Source code for metagrating topology optimization," GitHub (2023), <https://github.com/jonfanlab/Metagrating-Topology-Optimization>.
102. J. Jiang, R. Lupo, E. W. Wang, *et al.*, "Metanet: a new paradigm for data sharing in photonics research," *Opt. Express* **28**, 13670–13681 (2020).
103. H. Men, K. Y. Lee, R. M. Freund, *et al.*, "Robust topology optimization of three-dimensional photonic-crystal band-gap structures," *Opt. Express* **22**, 22632–22648 (2014).
104. S. Kim, T. Christensen, S. G. Johnson, *et al.*, "Automated discovery and optimization of 3D topological photonic crystals," *ACS Photonics* **10**, 861–874 (2023).
105. Z. Lin, L. Christakis, Y. Li, *et al.*, "Topology-optimized dual-polarization Dirac cones," *Phys. Rev. B* **97**, 081408 (2018).
106. F. Wang, J. S. Jensen, and O. Sigmund, "High-performance slow light photonic crystal waveguides with topology optimized or circular-hole based material layouts," *Photonics Nanostruct. Fundam. Appl.* **10**, 378–388 (2012).
107. M. L. Povinelli, S. G. Johnson, and J. D. Joannopoulos, "Slow-light, band-edge waveguides for tunable time delays," *Opt. Express* **13**, 7145–7159 (2005).

Validation and characterization of algorithms and software for photonics inverse design: supplement

MO CHEN,¹  RASMUS E. CHRISTIANSEN,^{2,3}  JONATHAN A. FAN,⁴  GÖKTUĞ IŞIKLAR,^{2,3}  JIAQI JIANG,⁴  STEVEN G. JOHNSON,^{1,*}  WENCHAO MA,⁵ OWEN D. MILLER,⁶  ARDAVAN OSKOOL,⁷  MARTIN F. SCHUBERT,^{8,9}  FENGWEN WANG,³  IAN A. D. WILLIAMSON,⁸  WENJIN XUE,⁶ AND YOU ZHOU⁴ 

¹Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

²NanoPhoton, Center for Nanophotonics, Technical University of Denmark, Ørsteds Plads 345A, 2800 Kongens Lyngby, Denmark

³Department of Civil and Mechanical Engineering, Solid Mechanics, Technical University of Denmark, DK-2800 Lyngby, Denmark

⁴Department of Electrical Engineering, Stanford University, Stanford, California 94305, USA

⁵Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

⁶Department of Applied Physics and Energy Sciences Institute, Yale University, New Haven, Connecticut 06511, USA

⁷Google, 1600 Amphitheatre Parkway, Mountain View, California 94043, USA

⁸X the moonshot factory, 100 Mayfield Avenue, Mountain View, California 94043, USA

⁹Current address: Meta, 1 Hacker Way, Menlo Park, California 94025, USA

*stevenj@math.mit.edu

This supplement published with Optica Publishing Group on 31 January 2024 by The Authors under the terms of the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/) in the format provided by the authors and unedited. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

Supplement DOI: <https://doi.org/10.6084/m9.figshare.24885930>

Parent Article DOI: <https://doi.org/10.1364/JOSAB.506412>

Supplementary Information:

Validation and characterization of algorithms and software for photonics inverse design

MO CHEN,¹ RASMUS E. CHRISTIANSEN,^{2,3} JONATHAN A. FAN,⁴
GÖKTUĞ IŞIKLAR,^{2,3} JIAQI JIANG,⁴ STEVEN G. JOHNSON,^{1,*}
WENCHAO MA,⁵ OWEN D. MILLER,⁶ ARDAVAN OSKOOI,⁷
MARTIN F. SCHUBERT,^{8,9} FENGWEN WANG,³
IAN A. D. WILLIAMSON,⁸ WENJIN XUE,⁶ AND YOU ZHOU⁴

¹Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

²NanoPhoton, Center for Nanophotonics, Technical University of Denmark, Ørsted Plads 345A, 2800 Kongens Lyngby, Denmark

³Department of Civil and Mechanical Engineering, Solid Mechanics, Technical University of Denmark, DK-2800 Lyngby, Denmark

⁴Department of Electrical Engineering, Stanford University, Stanford, California, 94305, USA

⁵Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

⁶Department of Applied Physics and Energy Sciences Institute, Yale University, New Haven, Connecticut 06511, USA

⁷Google, 1600 Amphitheatre Parkway, Mountain View, California 94043, USA

⁸X the moonshot factory, 100 Mayfield Avenue, Mountain View, California 94043, USA

⁹now at Meta, 1 Hacker Way, Menlo Park, California 94025, USA

*stevenj@math.mit.edu

1. Implementation Details for Estimation of Minimum Lengthscales

1.1. Structuring element

The structuring element or kernel, which can be understood as a probe, is a disc defined as $x^2 + y^2 \leq d^2/4$, where d is the diameter. Some probes with different diameters are shown in Fig. 1, where the size of a pixel is 1 along both x and y directions. Due to the finite resolution, the shapes of probes with small diameters deviate from a standard disc. As the diameter increases, the shape starts to resemble a standard disc.

1.2. Interior of Solid Regions

The interior of solid regions is obtained by applying morphological erosion using a small probe like a cross, as shown in Fig. 2.

1.3. Representation of Binary Images

The input image and the output of morphological operations are always converted to Boolean arrays. Solid and void pixels have values true and false, respectively. The operations in Eqs. (1–3)

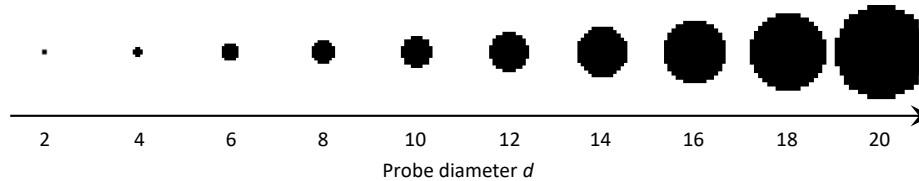


Fig. 1. Structuring elements or kernels with different diameters.

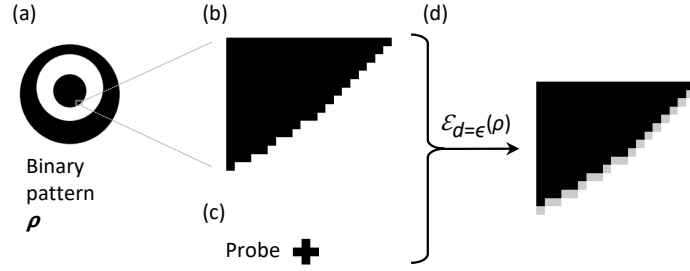


Fig. 2. Morphological erosion for extracting the interior of solid regions in a binary pattern. (a) Example of a binary design pattern. This pattern is the same as that in Fig. 2(a) of the main text. (b) Magnified image of the pattern in the box in (a). (c) Probe like a cross for morphological erosion. (d) Pattern after morphological erosion. The original pattern is superimposed as a guide to the eye.

of the main text, which compute the minimum solid/void/combined lengthscales, are implemented as element-wise Boolean operations. Specifically, multiplication, addition, and subtraction correspond to logical and, or, and exclusive or, respectively. Interchanging solid and void regions in a binary image is performed as negation of the Boolean array.

2. Design Details

2.1. RGB metalens

During each iteration of the FDTD optimization, the design pattern is blurred using a conic filter and then projected with increasing projection strength. The conic filtering can be described in the form of a convolution integral as

$$\tilde{\rho}(\mathbf{x}) = \int w(\mathbf{x} - \mathbf{y})\rho(\mathbf{y}), \quad (1)$$

with

$$w(\mathbf{x}) = \begin{cases} \frac{1}{w_n} \left(1 - \frac{|\mathbf{x}|}{R}\right), & |\mathbf{x}| < R, \\ 0, & |\mathbf{x}| \geq R, \end{cases} \quad (2)$$

where the normalization factor w_n needs to guarantee $\int w(\mathbf{x}) = 1$, and the filter radius R was inferred from the lengthscale [1] with $\eta_e = 0.6$ for this optimization problem. The projection of $\tilde{\rho}$ is defined as

$$\hat{\rho} = \text{Proj}(\tilde{\rho}) = \frac{\tanh(\beta\eta) + \tanh[\beta(\tilde{\rho} - \eta)]}{\tanh(\beta\eta) + \tanh[\beta(1 - \eta)]}, \quad (3)$$

where the threshold η is 0.5 while the projection strength β increases during optimization, e.g., starting with $\beta = 8$ and doubling every epoch with 100 iterations per epoch. The lengthscale constraints were only imposed for the last several epochs when the structures were fairly binarized. A typical evolution history is shown in Fig. 3.

The final design is exported as a CSV file and imported into another solver for validation. The FOMs (average enhancement of $|E|^2$ due to the metalens relative to air over three frequencies) from the optimization and subsequent validation using a different solver are provided in the Github repository https://github.com/NanoComp/photronics-opt-testbed/tree/main/RGB_metalens.

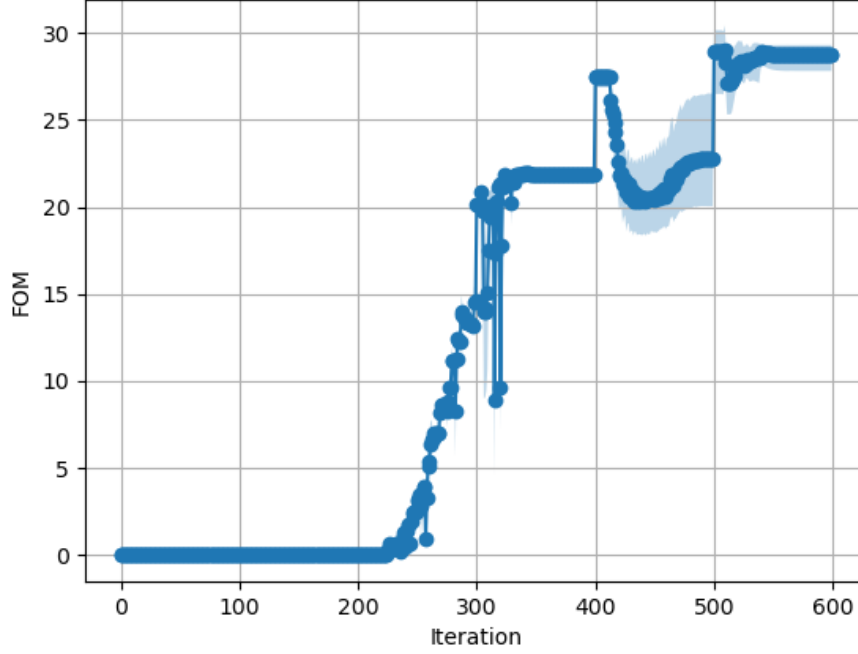


Fig. 3. Typical evaluation history of $|E|^2$ for RGB metalens. A lengthscale constraint is imposed only in the last two epochs (200 iterations) when the structure is binarized. This figure is for illustrative purposes as the vertical axis represents unnormalized values.

2.2. Metallic Concentrator

The metallic concentrators are designed using density-based topology optimization [2] applying filtering and thresholding developed in the mechanics community [3] to ensure a final binarized design. A filter radius of $r_f = 10$ nm is used for the filtering. For the thresholding a level of $\eta = 0.5$ is used with an initial thresholding strength of $\beta_{\text{initial}} = 8$. The thresholding strength is gradually increased during the optimization to a final value of 32. The initial guess for the inverse design process is taken to be $\rho = 1$ in the full design domain.

The objective function to be maximized is taken to be,

$$\Phi(\hat{\rho}) = |\mathbf{E}(\hat{\rho}, \mathbf{r}_0)|^2, \quad (4)$$

with the electric field $\mathbf{E}(\hat{\rho}, \mathbf{r}_0)$ being the solution to the time-harmonic Maxwell equations, Eq. (6) of the main text, over the model domain Ω for a given design field $\hat{\rho}$ and \mathbf{r}_0 being the target point.

The model domain, Ω , used in the design process is a square with a side length of $L_\Omega = 600$ nm. It is surrounded on all sides by a perfectly matched layer [4]. The design domain is a 2d torus with inner radius of 10 nm and outer radius of 100 nm, with the target point at its center. For the numerical discretization of the model problem the model domain is discretized using a triangular mesh. The design domain is discretized using elements with a minimum side length of 1 nm, dictating the resolution of the design, while the surrounding region is discretized using a minimum of 16 elements per free-space wavelength to ensure accurate modelling. First

order finite element basis functions are used as trial and test functions for discretizing the model equations.

The wavelength of the light in the model problem is taken to be $\lambda = 532$ nm and the material parameters for the metal are taken from [5].

The optimization problem is solved using the globally convergent method of moving asymptotes (GCMMA) [6] with a minimum of 70 design iterations per β -value and the stopping criterion being that the objective function value did not change more than 0.1% over ten consecutive iterations.

The final design was post processed by sampling it on a 0.1 nm grid and filtering the sampled design using a 1.5 nm filter radius from which the final design was extracted as the 0.5-contour of the filtered field. This post-processed design geometry was then evaluated on a body fitted triangular mesh with elements of a 1 nm side length along the surface of the design and a minimum of 16 elements per free-space wavelength away from the design, using quadratic Lagrange basis functions for high accuracy in the final evaluation.

2.3. Waveguide Mode Converter

Exemplary waveguide mode-converter designs were generated by two distinct schemes. The first method is described in [7] and uses a straight-through estimator (STE) along with a conditional generator for designs having a prescribed minimum lengthscale. Combined, these constitute a differentiable transform; the transform is applied to a latent design, which is the underlying optimization variable of the problem. The latent design is optimized by the first-order Adam algorithm, with learning rate of 0.01 and hyper parameters $\beta_1 = 0.667$ and $\beta_2 = 0.9$. Designs are simulated using the 2D FDFD simulator Ceviche, and the objective function given in [7] is used, which targets reflection less than -20dB and transmission greater than -0.5dB across all wavelengths. For each point in Fig. 5 of the main text (the mode-converter results), a randomly-initialized optimization is carried out for 1000 steps, and the best result is reported. Hyper parameters are not adjusted during the course of the optimization. The cumulative minimum loss versus step for several different lengthscale targets is shown in Fig. 4. The scalar-valued objective function is [7]:

$$\Phi(S_{11,\lambda}, S_{21,\lambda}) = \sum_{\lambda} \text{softplus} \left(\frac{T_{\text{cutoff}} - |S_{21,\lambda}|^2}{1 - T_{\text{cutoff}}} \right)^2 + \sum_{\lambda} \text{softplus} \left(\frac{|S_{11,\lambda}|^2 - R_{\text{cutoff}}}{R_{\text{cutoff}}} \right)^2, \quad (5)$$

where $T_{\text{cutoff}} = 10^{-\frac{0.5}{10}} \approx 0.891$ and $R_{\text{cutoff}} = 10^{-\frac{20}{10}} = 0.01$.

The second method is described in [8] and involves 2D FDTD simulations using Meep to perform worst-case (minimax) optimization whereby the objective function is based on minimizing the maximum of (1) reflection into the first-order mode of the input port and (2) the transmission into the second-order mode of the output port across six wavelengths. The minimax optimization is implemented using an epigraph variable method using the method of moving asymptotes (MMA) algorithm of the NLOpt package [9]. The optimization involved six epochs in which the bias parameter of the projection operator was repeatedly doubled from 8 to 256. The initial design of the first epoch used a constant value of 0.5 for all weights. Subsequent epochs used the optimal design of the previous epoch as the initial design. There are four important details of the optimization procedure: (1) the minimum linewidth constraint of [1] is activated only in the final epoch, (2) the edge of the design region is padded by a conical filter radius to prevent violations of the linewidth constraint near the edge of the design region, (3) hyperparameters of the linewidth constraint are not adjusted during the course of the optimization, and (4) subpixel smoothing of the design weights is used when the bias parameter is larger than 64 for which the design is nearly binarized.

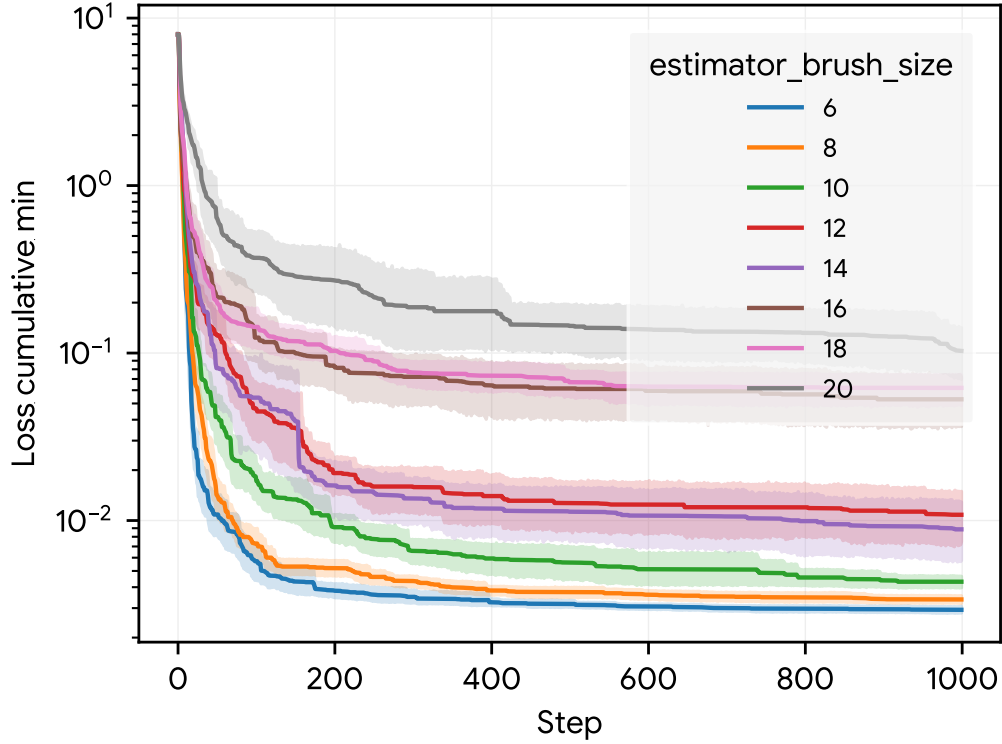


Fig. 4. Cumulative minimum loss for optimization of the waveguide mode converter for several different lengthscale targets. The estimator brush size is in units of 10 nm pixels. The shaded regions indicate the range of cumulative minimum loss value across different randomly seeded optimizations and the solid line represents the mean cumulative minimum loss across the different optimizations. The minimum operation is applied per optimization cumulatively across optimization step.

2.4. Cavity Design

Similar to the RGB metalens problem, to achieve better results, several hyper-parameters were fine-tuned for each case individually. Generally, the raw design parameters were conic-filtered with filter radius around the target feature size, and then tanh projected with β gradually increased throughout the optimization. We noted that the performance of cavities are extremely sensitive to small changes, so β has to be ramped up slowly to have less changes at each step; on the other hand, the optimization tends to converge quickly at each β . As a result, we started with reasonably large β (e.g. $\beta = 24$) and employed scheme $\beta \rightarrow 1.2\beta$ every 25 epochs. The lengthscales constraints were only imposed during the last several epochs when the structures were fairly binarized. Instead of directly maximizing LDOS, we also found minimizing $\frac{1}{\text{LDOS}}$ to be beneficial. A typical evolution history is shown in Fig 6.

After final designs were obtained, they were exported as binary .csv files and imported to another solver for validation. Since different solvers have different interpolations onto the respective discretization, we first found the exact resonant frequency for each structure in that solver, which may differ slightly from the original target, and then calculated the LDOS at that resonant frequency.

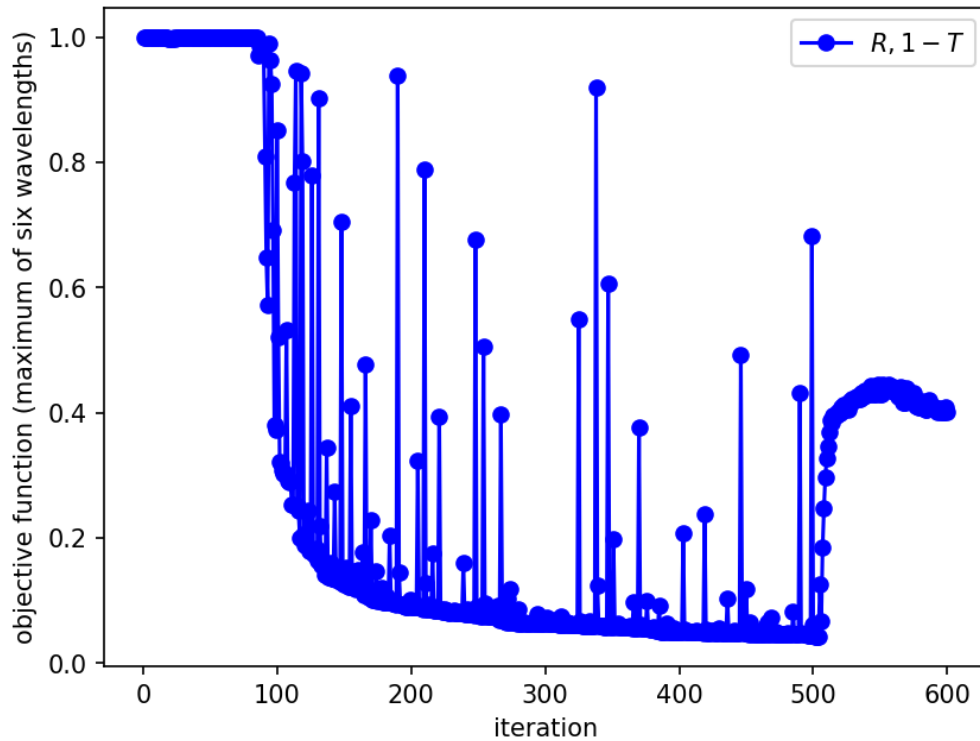


Fig. 5. Objective-function history (the maximum of the reflectance and 1-transmittance for six wavelengths) for the worst-case optimization of a waveguide mode converter with a minimum feature size of 150 nm using 2D FDTD simulations via Meep. The "spikes" present in the plot are a normal feature of nonlinear-optimization algorithms. The algorithm may take too large a step which turns out to make the objective function worse. This means the algorithm then has to "backtrack" and take a smaller step. This occurs in the CCSA algorithm by increasing a penalty term. Also, even though the worst-case objective function is constant during most of the first epoch which may indicate the optimizer is making no progress, in fact the optimizer is working to improve the objective function at the other (non worst-case) wavelengths.

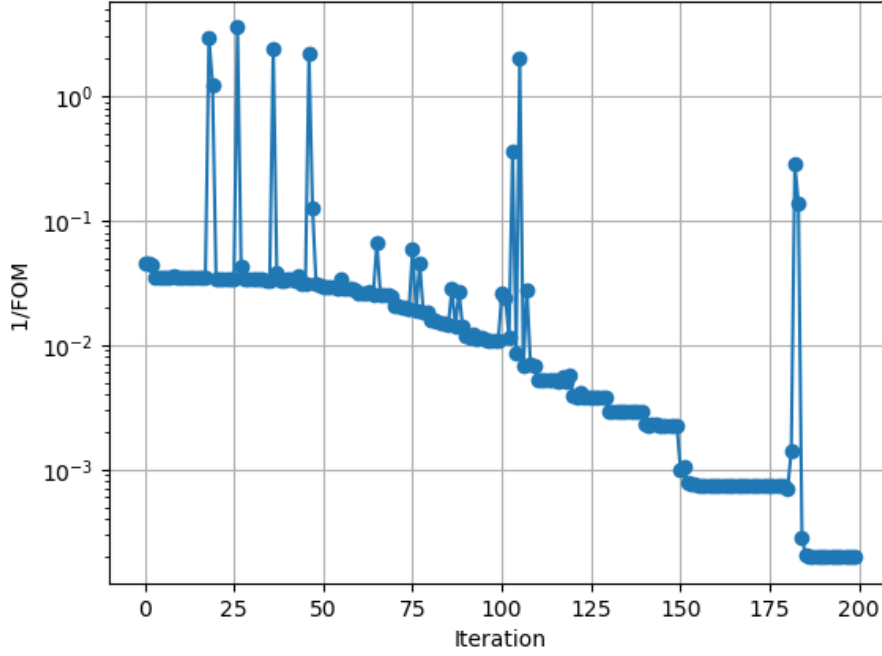


Fig. 6. Typical evaluation history of $\frac{1}{\text{LDOS}}$ for cavity design. lengthscale constraint was imposed in the last 25 iterations. The figure is for illustrative purpose, as the vertical axis represents unnormalized values.

2.5. 3D Metagrating

In each iteration of the FDTD optimization, the design pattern were blurred by a conic filter with a filter radius equal to 60 nm and then projected with the threshold 0.5 with increasing projection strength $\beta = 4, 8, 16, 32, 64, 96, 128, 256, \infty$. When the optimization with a finite value of β undergoes 80 iterations or satisfies the stopping criterion, β is increased to the next value. The evaluation history of the diffraction efficiency in optimization is in Fig. 7.

All designs were performed with the resolution of 90 points per wavelength. In validation, the simulation resolution was 360 points per wavelength and the arrays representing the designs were also interpolated to this resolution using bicubic interpolation. The FOMs (diffraction efficiencies at $\lambda = 1050$ nm) in validation are summarized in Table 1.

Lengthscale (nm)	FOM from RCWA	FOM from FDTD
66	93.2%	93.8%
62	96.8%	95.0%
54	93.7%	92.5%
35	83.0%	84.3%

Table 1. Lengthscales and FOMs of the 3d metagratings.

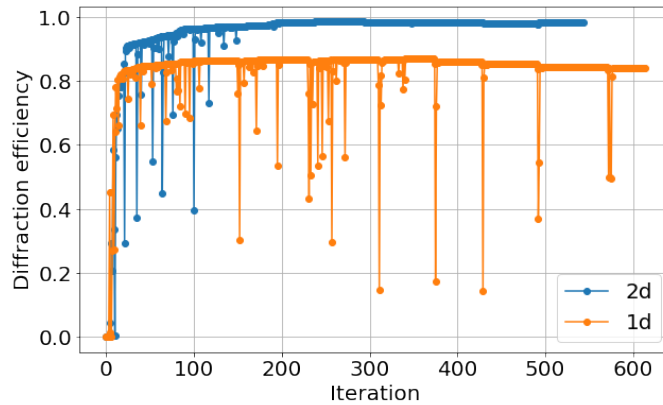


Fig. 7. Evaluation history of diffraction efficiency in designing 3d metagratings based on FDTD. The blue and orange curves correspond to the 2d design in Fig. 8(d) of the main text and the 1d design in Fig. 8(e) of the main text, respectively.

2.6. Dispersion Engineering

In each optimization iteration, there are in average around 3-5 inner objective evaluations for GCMMA. Fig. 8 shows the optimization history of the dispersion engineering problem, i.e., objective v.s. function evaluation number.

References

1. A. M. Hammond, A. Oskooi, S. G. Johnson, and S. E. Ralph, "Photonic topology optimization with semiconductor-foundry design-rule constraints," *Opt. Express* **29**, 23916–23938 (2021).
2. R. E. Christiansen and O. Sigmund, "Inverse design in photonics by topology optimization: tutorial," *JOSA B* **38**, 496–509 (2021).
3. F. Wang, B. S. Lazarov, and O. Sigmund, "On projection methods, convergence and robust formulations in topology optimization," *Struct. Multidiscip. Optim.* **43**, 767–784 (2011).
4. J.-P. Berenger, "A perfectly matched layer for the absorption of electromagnetic waves," *J. Comput. Phys.* **114**, 185–200 (1994).
5. P. B. Johnson and R. W. Christy, "Optical constants of noble metals," *Phys. Rev. B* **6(12)**, 4370–4379 (1972).
6. K. Svanberg, "A class of globally convergent optimization methods based on conservative convex separable approximations," *SIAM journal on optimization* **12**, 555–573 (2002).
7. M. F. Schubert, A. K. C. Cheung, I. A. D. Williamson, A. Spyra, and D. H. Alexander, "Inverse design of photonic devices with strict foundry fabrication constraints," *ACS Photonics* **9**, 2327–2336 (2022).
8. A. M. Hammond, A. Oskooi, M. Chen, Z. Lin, S. G. Johnson, and S. E. Ralph, "High-performance hybrid time/frequency-domain topology optimization for large-scale photonics inverse design," *Opt. Express* **30**, 4467 (2022).
9. "The NLOpt nonlinear-optimization package," <http://github.com/stevengj/nlopt>.

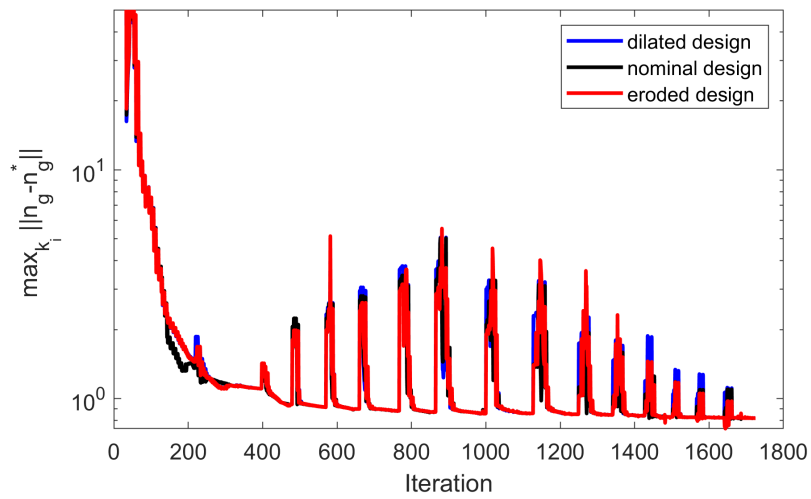


Fig. 8. Evaluation history of dispersion engineering for the slow light waveguide design using the three-case robust formulation. Blue: dilated design generated using $\eta = 0.35$; Black: optimized design generated using $\eta = 0.5$; Red: eroded design generated using $\eta = 0.65$; The plot shows the largest error between the actual and prescribed group index in the target wavenumber interval for the three considered design realizations in all the function evaluations during optimization. The low bounds represent the real optimization history using GCMMA, and the "spikes" present in the plot occur in the linear search part inside each GCMMA optimization iteration.