

MIT Open Access Articles

*Photonic topology optimization with
semiconductor-foundry design-rule constraints*

The MIT Faculty has made this article openly available. **Please share**
how this access benefits you. Your story matters.

Citation: Hammond, Alec et al. "Photonic topology optimization with semiconductor-foundry design-rule constraints." *Optics Express* 29, 15 (June 2021): 23916-23938. © 2021 Optical Society of America

As Published: <http://dx.doi.org/10.1364/oe.431188>

Publisher: The Optical Society

Persistent URL: <https://hdl.handle.net/1721.1/131097>

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of Use: Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.



Photonic topology optimization with semiconductor-foundry design-rule constraints

ALEC M. HAMMOND¹, ARDAVAN OSKOOI^{2,†}, STEVEN G. JOHNSON³,
STEPHEN E. RALPH^{1*}

¹*School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30308, USA*

²*Simpetus, LLC, San Francisco, CA 94109, USA*

³*Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

[†]*now at X, the Moonshot Factory, Mountain View, CA 94043, USA*

*stephen.ralph@ece.gatech.edu

Abstract: We present a unified density-based topology-optimization framework that yields integrated photonic designs optimized for manufacturing constraints including all those of commercial semiconductor foundries. We introduce a new method to impose minimum-area and minimum-enclosed-area constraints, and simultaneously adapt previous techniques for minimum linewidth, linespacing, and curvature, all of which are implemented without any additional re-parameterizations. Furthermore, we show how differentiable morphological transforms can be used to produce devices that are robust to over/under-etching while also satisfying manufacturing constraints. We demonstrate our methodology by designing three broadband silicon-photonics devices for nine different foundry-constraint combinations.

© 2021 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

Photonic topology optimization (TO) or “inverse design” is a powerful device-design methodology in which performance (e.g. transmitted power) can be optimized over thousands or millions of degrees of freedom characterizing every “pixel” of the fabricated device, allowing completely unexpected designs with unprecedented performance to emerge [1–11]. The vast design freedom of TO must be constrained, however, by the limitations of practical manufacturing processes, and in previous TO work this has taken the form of various techniques to constrain the minimum lengthscales and curvature present in the optimized design [12–15]. In order to be useful within the commercial semiconductor foundry ecosystem, however, additional design constraints must be developed [16, 17]: foundry lithography requires a minimum area and a minimum enclosed area, in addition to minimum curvature and lengthscales (linewidths and linespacings). The theoretical challenge is to formulate these constraints in a manner that is mathematically compatible with practical TO algorithms.

In this paper, we present a unified framework for density-based topology optimization which produces integrated photonic devices that obey complicated commercial foundry design rule checks (DRC). To begin with, we consolidate previous techniques to impose minimum linewidth, linespacing, and curvature constraints [13, 18, 19] (Sec. 3). In addition, we present a new TO-compatible formulation of minimum-area and minimum enclosed-area constraints (Sec. 4). Our methodology relies only on classical density-based projection methods [20] and does not require any additional reparameterization [12, 21–24] (e.g. level sets or splines). Furthermore, we combine our manufacturing constraints with *robust* optimization [14, 25, 26], incorporating under/over-etch uncertainty into the TO design process (Sec. 5), whereas previous work typically implemented robustness without manufacturing constraints [25] or vice-versa. We demonstrate our approach on three broadband silicon-photonics test problems (Sec. 6)—dielectric waveguide bends, T-splitters, and reflectors—all optimized for broadband operation using time-domain topology optimization [27, 28] developed using the free/open-source software package Meep [29].

Density-based TO methods, so named because the design parameters directly correspond to the local material properties, typically require a sequence of standard image-processing steps, such as filtering and thresholding, to map the design parameters to the physical structure, allowing the optimization algorithm to act on a differentiable function (Sec. 2) [30]. The sensitivity of this resulting structure is efficiently calculated using an adjoint-variable method requiring just two electromagnetic simulations, regardless of the number of design variables [31]. In the time domain, a short pulse can excite a broadband response that is then used to calculate the respective sensitivities at multiple frequency points simultaneously [27, 28]. Since the optimizer is gradient-based, manufacturing constraints need to be formulated in the form $g(\text{design}) \leq 0$ where g is some function that is (mostly) differentiable. Our approach follows this general framework exactly, allowing it to be directly incorporated into pre-existing TO schemes without any need to reparameterize [32] during optimization, effectively automating the design process for fabrication.

Ensuring that a final design is “DRC clean” is standard practice in the foundry ecosystem, where thousands of checks are performed to verify that the geometries can be reliably fabricated and pose no risk to surrounding designs or the manufacturing tooling [16]. While circuit-level DRC violations can be algorithmically fixed using design post-processing [33], devices designed with TO must integrate these design rules into the optimization process itself. Important “1D” constraints such as the minimum linewidth, linespacing, and curvature are well understood within the context of TO [12, 13]. More complicated “2D” constraints, like minimum areas and enclosed areas, require new developments, especially since these constraints are often more restrictive than their 1D counterparts and currently are only enforceable through static B-spline parameterizations [23], limiting the problem to shape-optimization methods that require a pre-determined material topology (connectivity, number of holes, etcetera). Fig. 1 illustrates each of these fundamental design rules on three example problems produced using TO. Indeed many other rules, including the maximum area, run length, and prescribed density [16] have also been overlooked in TO. Our experiments demonstrate that satisfying the five aforementioned fundamental constraints, however, often translates implicitly to satisfying these “higher order” constraints as well.

Even after DRC violations are cleared, integrated photonic devices may exhibit sensitivity to even minor process variations. Robust optimization generally means optimizing the worst case of an objective over some set of uncertainties [34]. Correspondingly, in Sec. 5 we optimize the worst of the design, an under-etched design, and an over-etched design [34]. In contrast to previous robust TO [14, 25, 26, 34], our approach decouples the DRC constraints from over/under-etch uncertainties, meaning that the topologies between the eroded, dilated, and blueprint design fields need no longer remain consistent. Similarly, the erosions and dilations can now be applied by a simple formula, independent of the dynamic range of the underlying design field. Our methodology, validated by dozens of numerical experiments on nine unique foundry rulebooks, provides a simple and cohesive framework for practical photonic TO targeting high-yield foundry applications.

2. Photonic topology optimization overview

In this section, we review photonics TO, following Ref. 20, in the context of semiconductor foundry design and discuss various projection methods. We first formulate broadband design as a minimax optimization problem over N distinct objective functions subject to Maxwell’s

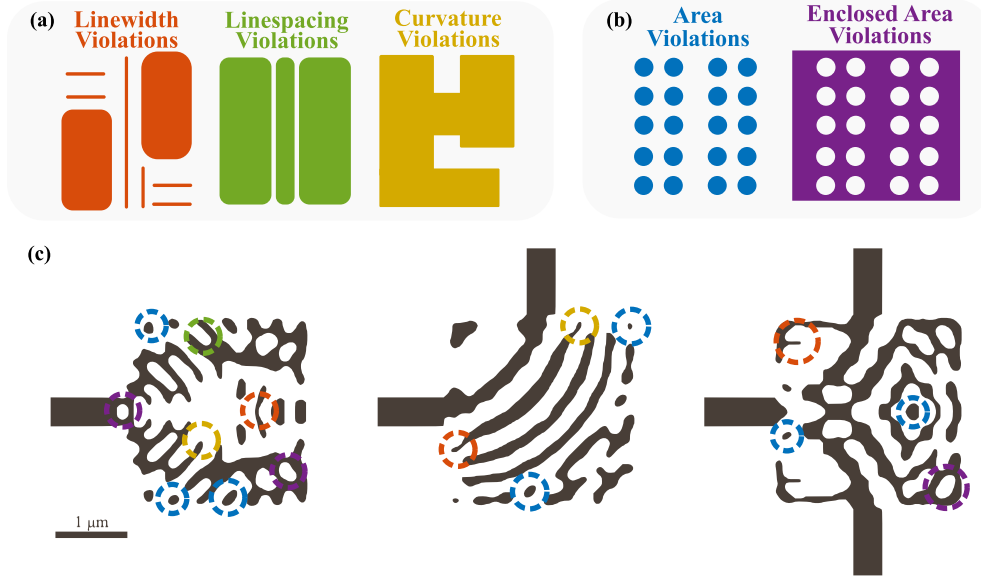


Fig. 1. Five fundamental design rule constraints for semiconductor foundries. 1D constraints consist of minimum linewidth, minimum linespacing and minimum curvature (a). 2D constraints include minimum area and minimum enclosed area (b). Example DRC violations in a broadband mirror, broadband bend, and broadband T-splitter are highlighted in respective colors (c).

equations spanning M frequency points:

$$\begin{aligned}
 & \min_{\rho} [\max_n \{f_n(\mathbf{E})\}] & n \in \{1, 2, \dots, N\} \\
 \text{s.t. } & \nabla \times \frac{1}{\mu_0 \mu_r} \nabla \times \mathbf{E} - \omega_m^2 \epsilon_0 \epsilon_r(\boldsymbol{\rho}) \mathbf{E} = -i\omega_m \mathbf{J} & m \in \{1, 2, \dots, M\} \\
 & 0 \leq \rho \leq 1 \\
 & g_k(\boldsymbol{\rho}) \leq 0 & k \in \{1, 2, \dots, K\}
 \end{aligned} \tag{1}$$

where f_n is each objective function dependent on the field pattern \mathbf{E} , ϵ is the relative permittivity as a function of the density design variables $\boldsymbol{\rho}$ at each point in space, \mathbf{J} is the current density, and g_k is the k^{th} constraint function. We note that for the majority of this work, $N = M$, implying that each objective function corresponds to a unique frequency point, and each frequency point uses the same objective function, although the method is generalizable to multiple (same or different) objective functions per frequency point (e.g. for robust optimization when each frequency point has multiple design fields). By using an epigraph formulation [35], we can recast the non-differentiable minimax problem above into a differentiable minimization problem with additional nonlinear constraints:

$$\begin{aligned}
 & \min_{\rho, t} t \\
 \text{s.t. } & \nabla \times \frac{1}{\mu_0 \mu_r} \nabla \times \mathbf{E} - \omega_m^2 \epsilon_0 \epsilon_r(\boldsymbol{\rho}) \mathbf{E} = -i\omega_m \mathbf{J} & m \in \{1, 2, \dots, M\} \\
 & 0 \leq \rho \leq 1 \\
 & f_n(\mathbf{x}) - t \leq 0 & n \in \{1, 2, \dots, N\} \\
 & g_k \leq 0 & k \in \{1, 2, \dots, K\}
 \end{aligned} \tag{2}$$

where $t \in \mathbb{R}$ is a dummy parameter.

In order to parameterize the permittivity $\epsilon_r(\boldsymbol{\rho})$, we use a simple density-based interpolation scheme where the original (latent) design parameters $\boldsymbol{\rho}$, are first filtered using a linear density filter,

$$\tilde{\boldsymbol{\rho}} = w(\mathbf{x}) * \boldsymbol{\rho}, \quad (3)$$

$\tilde{\boldsymbol{\rho}}$ is the filtered design field, w is the filter kernel, and $*$ is 2D convolution [20]. Our methodology's design rule constraints only require two different filter kernels. The first is commonly referred to as the "uniform" or "top-hat" kernel and is defined by

$$w(\mathbf{x}) = \begin{cases} \frac{1}{|\mathcal{N}|} & \mathbf{x} \in \mathcal{N} \\ 0 & \mathbf{x} \notin \mathcal{N} \end{cases}, \quad (4)$$

where the filter support itself is described by the domain, \mathcal{N} , and used to smooth the design variables uniformly. In the case of foundry-based optimization, which requires all designs to be 3D extrusions of 2D shapes, common choices for \mathcal{N} include a circle, square, or ellipse.

The second filter kernel is a conic filter, as the strength of the filter linearly decays toward the edge of the filter support and is defined by

$$w(\mathbf{x}) = \begin{cases} \frac{1}{a} \left(1 - \frac{|\mathbf{x} - \mathbf{x}_0|}{R}\right) & \mathbf{x} \in \mathcal{N} \\ 0 & \mathbf{x} \notin \mathcal{N} \end{cases} \quad (5)$$

where \mathcal{N} is a circle of radius R , \mathbf{x}_0 is the center of \mathcal{N} , and a is a normalization factor such that $\int w(\mathbf{x}) = 1$.

Once the design variables are filtered using the appropriate kernel, either Eq. (4) or Eq. (5), the resultant field is projected onto a binary value using a differentiable and nonlinear function. In this work, we use

$$\bar{\boldsymbol{\rho}} = \frac{\tanh(\beta\eta) + \tanh(\beta(\tilde{\boldsymbol{\rho}} - \eta))}{\tanh(\beta\eta) + \tanh(\beta(1 - \eta))} \quad (6)$$

where $\bar{\boldsymbol{\rho}}$ is the projected design field, and β and η are the threshold parameters [36]. Fig. 2 illustrates the effect of different filter kernels on the same set of design parameters, along with the resulting projection.

The final permittivity is then interpolated from the design parameters using

$$\epsilon_r(\bar{\boldsymbol{\rho}}) = \epsilon_{\min} + \bar{\boldsymbol{\rho}}(\epsilon_{\max} - \epsilon_{\min}) \quad (7)$$

where ϵ_{\min} is the permittivity of the "void" region (cladding) and ϵ_{\max} is the permittivity of the "solid" region (core).

Using this methodology, we can systematically erode and dilate the design field using two different approaches, as illustrated in Fig. 3. The first approach adjusts the value of η used in the projection step. Dilation is produced by $\eta < 0.5$ and erosion is produced by $\eta > 0.5$. While often used in practice [14], the resulting erosions and dilations are directly dependent on the rate of change of the filtered design field $\tilde{\boldsymbol{\rho}}$. In other words, using the same value for η on two different filtered profiles does not guarantee the same amount of erosion/dilation. This is inconvenient when trying to robustly design for a specific amount of over- and under-etching, as is further discussed in Sec. 5.

To overcome this issue, we propose a new approach which assumes $\eta = 0.5$ for all elements of $\boldsymbol{\rho}$, but involves a nonlinear filter step *after* the projection. The harmonic erosion filter [37] is defined by

$$\mathcal{E}_{\mathcal{N}}(\boldsymbol{\rho}) = \left(\frac{1}{\boldsymbol{\rho} + \alpha} * w \right)^{-1} - \alpha \quad (8)$$

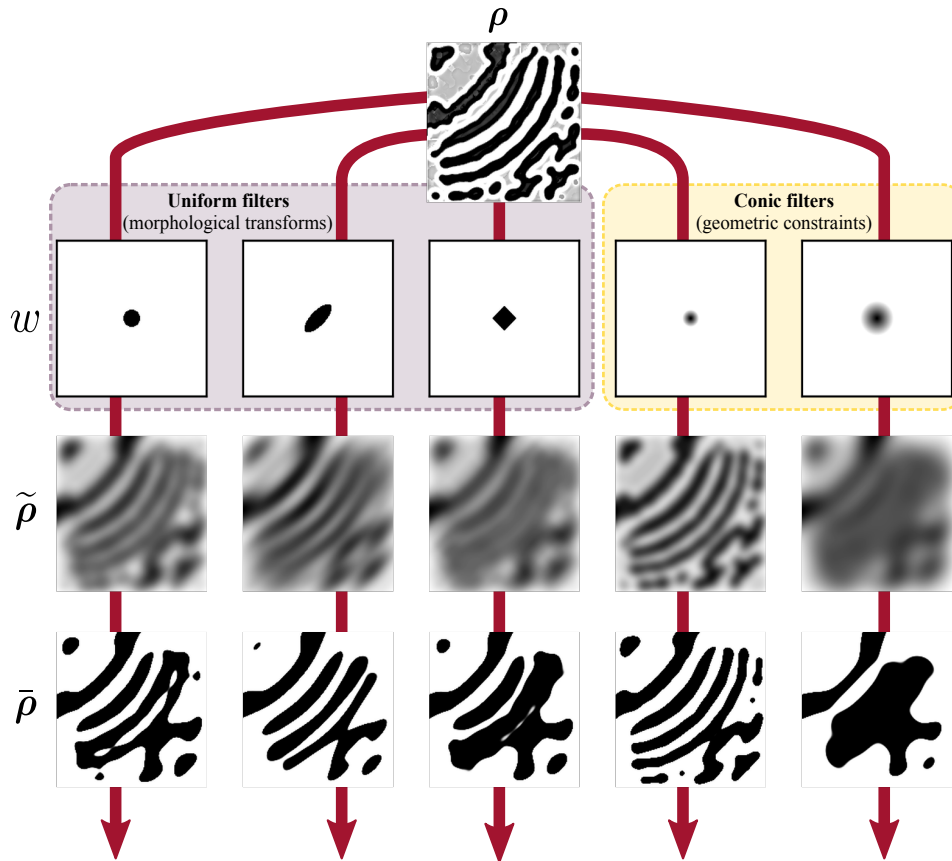


Fig. 2. Design variables, ρ , evolution as they are convolved with various filter kernels, w , forming a filtered field, $\tilde{\rho}$, and subsequently a projected field, $\bar{\rho}$. The filter kernels from left to right are a uniform cylinder, a uniform ellipse rotated by 45° , a uniform square rotated by 45° , a conic filter, and a conic filter with twice the radius. The uniform filter kernels (marked in pink) are used for morphological transforms (Sec. 3.2) that enable systematic robust optimization (Sec. 5). The conic filters (marked in yellow) are used for the geometric constraints (Sec. 3.1) that enforce minimum linewidth and linespacing constraints.

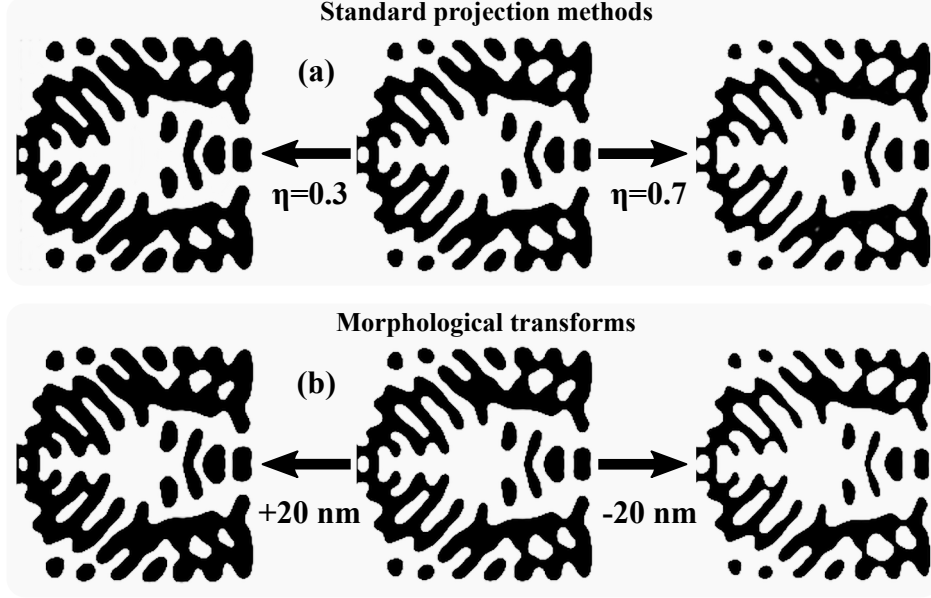


Fig. 3. Comparison of dilated (left) and eroded (right) design fields using the standard projection method (a) and harmonic filters (b). The harmonic filters can explicitly perform uniform erosions and dilations on the original design field, but are limited by the design field resolution and require an extra nonlinear filter step. In contrast, the standard projection method has more flexibility at the expense of predictability. The erosions and dilations are uniform for each contour and depend on the underlying design variable distribution.

where α is a nonlinear threshold parameter. The harmonic dilation is defined by

$$\mathcal{D}_N(\rho) = 1 - \left(\frac{1}{1 - \rho + \alpha} * w \right)^{-1} - \alpha \quad (9)$$

The filter w is typically a uniform kernel and its support determines the amount of erosion/dilation that occurs in a particular direction. α is a regularization parameter to ensure differentiability; the exact dilation/erosion operation is obtained in the $\alpha \rightarrow 0$ limit, and we find that $\alpha = 10^{-3}$ is sufficient for usage in TO. For example, a filter support defined by a circle with a radius of 20 nm will systematically erode/dilate the design by 20 nm uniformly in all directions. The advantage of this approach is that it performs the same morphological transformation regardless of the rate of change of the underlying filtered field, as the nonlinear filters are applied after the projection step (and the resulting structures are already quasi-binary).

3. Minimum linewidth, spacing, and curvature constraints

Here, we consolidate existing methodologies for lengthscale constraints from several papers [13, 18, 26, 36] to ensure that our designs obey minimum-linewidth, minimum-linespacing, and minimum-curvature requirements.

3.1. Geometric constraints

Ref. 13 introduced two geometric constraints that enforce a minimum lengthscale on both the solid and void regions of a topology. These particular constraints correspond exactly to the

minimum-linewidth and minimum-linespacing constraints of a foundry’s design rulebook.

The constraints work by identifying inflection regions that dictate whether or not the design topology remains unchanged after a strategically chosen set of erosions and dilations. If the topology *is* consistent (i.e. no additional “holes” or “islands”), then the lengthscale is satisfied. Enforcing an optimization constraint that implements this process requires an indicator function capable of identifying the inflection regions and a known relationship between the filter kernel and the necessary erosions/dilations.

The minimum-linewidth constraint, $g_{LW} \leq 0$, is described by the function

$$g_{LW} = \frac{1}{n} \sum_{i \in \mathbb{N}} I_i^{LW}(\rho_i) \cdot [\min\{\tilde{\rho}_i - \eta_e, 0\}]^2 \quad (10)$$

where $I_i^{LW}(\rho)$ is an indicator function that identifies the inflection region of the solid phase:

$$I_i^{LW}(\rho) = \bar{\rho} \cdot \exp\left(-c |\nabla \tilde{\rho}|^2\right). \quad (11)$$

where c is a dampening term that dictates the “strength” of the indicator function. The minimum-linespacing constraint, $g_{LS} \leq 0$, is described by a similar function

$$g_{LS} = \frac{1}{n} \sum_{i \in \mathbb{N}} I_i^{LS}(\rho_i) \cdot [\min\{\eta_d - \tilde{\rho}_i, 0\}]^2 \quad (12)$$

where $I_i^{LS}(\rho)$ is an indicator function that identifies the inflection region of the void phase:

$$I_i^{LS}(\rho) = (1 - \bar{\rho}) \cdot \exp\left(-c |\nabla \tilde{\rho}|^2\right). \quad (13)$$

To determine η_e and η_d from the foundry minimum linewidth (l_w) and minimum spacing (l_s), the following relations hold [26]:

$$\eta_e = \begin{cases} \frac{1}{4} \left(\frac{l_w}{R}\right)^2 + \frac{1}{2}, & \frac{l_w}{R} \in [0, 1] \\ -\frac{1}{4} \left(\frac{l_w}{R}\right)^2 + \frac{l_w}{R}, & \frac{l_w}{R} \in [1, 2] \\ 1, & \frac{l_w}{R} \in [2, \infty) \end{cases} \quad (14)$$

$$\eta_d = \begin{cases} \frac{1}{2} - \frac{1}{4} \left(\frac{l_s}{R}\right)^2, & \frac{l_s}{R} \in [0, 1] \\ 1 + \frac{1}{4} \left(\frac{l_s}{R}\right)^2 - \frac{l_s}{R}, & \frac{l_s}{R} \in [1, 2] \\ 0, & \frac{l_s}{R} \in [2, \infty) \end{cases}, \quad (15)$$

where R is the user-specified radius of a conic filter. This allows one to arbitrarily choose the filter radius and derive the subsequent threshold parameters (η_e and η_d) from the above relations. Choosing smaller radii allows for smaller features at the expense of potentially “stiffening” the corresponding optimization problem (i.e., the constraint Hessian may have a larger condition number). Typical values for η_e and η_d are 0.75 and 0.25, respectively [11, 13], which corresponds to $R = 2l_w = 2l_s$. In the case where $l_w > l_s$, one possible solution is to use $\eta_e = 0.75$, $R = 2l_w$, while η_d is calculated from the relations above. If $l_w < l_s$, then one could use $\eta_d = 0.25$, $R = 2l_s$, while η_e is similarly calculated from (14). We use this approach for all of our numerical examples in Sec. 6.

For the dampening coefficient c , we use $c = r^4$ where r is the design grid resolution (i.e. $\frac{1}{\Delta x}$). As we discuss in Sec. 6, we impose the constraint gradually during optimization by employing a constraint $g_k \leq G_k$, where G_k is reduced in a systematic way as the optimization progresses. The

choice of tolerances G_k allows some flexibility in the tradeoff between the geometric constraints and the area constraints.

Finally, we note that both the linewidth and linespacing constraints also place implicit constraints on the corresponding minimum curvature for both solid and void regions. Assuming a circular filter is used and that the constraint is satisfied, the resulting topology cannot contain circular elements whose diameters are smaller than the corresponding lengthscale (e.g. linewidth for solid regions and linespacing for void regions). Consequently, the corresponding minimum radius of curvature $\kappa_{w,s}$ is defined by

$$\kappa_{w,s} = \frac{l_{w,s}}{2}. \quad (16)$$

In some cases it may be desirable to place a smaller radius of curvature constraint than is dictated by the minimum linewidth or spacing (e.g. for platforms that are able to resolve sharper angles). In such cases, a morphological transform may be a better approach.

3.2. Morphological-transform constraints

As an alternative to the geometric constraints described above, it has been proposed that one can enforce minimum linewidth, linespacing, and curvature constraints using differentiable morphological transforms [18]. In practice, we found these morphological constraints to be inconvenient for generating binary designs as described below. However, we review them here because we found them to be useful, instead, in implementing *robust* optimization as described in Sec. 5.

First, from the erosion and dilation operators of eqs. (8–9), one defines an open operator (\mathcal{O}_N)

$$\mathcal{O}_N(\rho) = \mathcal{D}_N(\mathcal{E}_N(\rho)), \quad (17)$$

which is simply an erosion followed by a dilation (performed using the same filter neighborhood, \mathcal{N}). Similarly, one defines the close operator (\mathcal{C}_N)

$$\mathcal{C}_{N'}(\rho) = \mathcal{E}_{N'}(\mathcal{D}_{N'}(\rho)) \quad (18)$$

which is simply a dilation followed by an erosion (using the filter neighborhood \mathcal{N}'). (Note that the open and close operators may have distinct filter neighborhoods $\mathcal{N} \neq \mathcal{N}'$.) If an open operator produces the same design field as a closed operator such that

$$\mathcal{O}_N(\rho) = \mathcal{C}_{N'}(\rho) \quad (19)$$

then the resulting design field must satisfy the minimum linespacing, linewidth, and curvature constraints as dictated by the filter kernels themselves. Specifically, the shape \mathcal{N} of the *open* operator's kernel determines the constraints on the void regions, and the shape \mathcal{N}' of the *close* operator determines the constraints on the solid regions. For example, a *close* operator with a rectangular kernel with side length of 90 nm and rounded corners with a radius of 10 nm would enforce a linewidth constraint of 90 nm and a radius of curvature of 10 nm. Thus, \mathcal{N} and \mathcal{N}' parameterize four distinct design rules.

While elegant, this approach still requires a suitable constraint function that is differentiable. One obvious implementation might be an L^2 norm

$$g_{LW,LS,\kappa} = \|\mathcal{O}_N(\rho) - \mathcal{C}_{N'}(\rho)\|_2 \leq G_{LW,LS,\kappa}, \quad (20)$$

for a threshold $G_{LW,LS,\kappa}$. We find that a practical challenge with this approach in practice is that the constraint does not encourage a truly binary design once enforced. Often, we observe that the optimizer will stall trying to satisfy the constraint if the design itself is not already sufficiently binarized. For this reason, we typically use the geometric constraints of Sec. 3.1, but we use the open/close operators for robust optimization in Sec. 5.

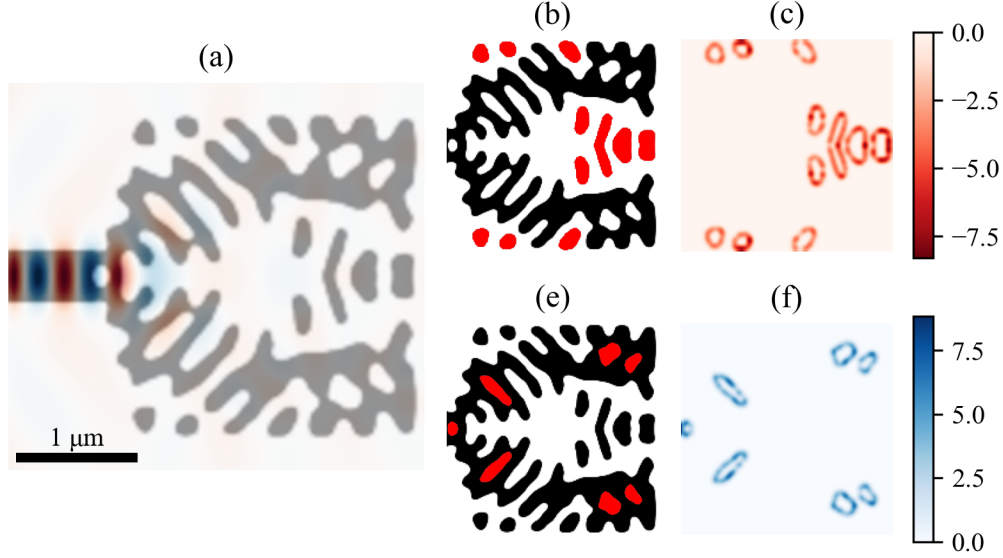


Fig. 4. Minimum area and minimum enclosed area constraint functions. (a) The nominal design under consideration is first parsed using an image segmentation algorithm to identify both the (b) islands and (e) holes that correspond to regions violating either constraint (highlighted in red). These contour masks serve as indicator functions for the optimization constraint. Gradients for the (c) minimum area constraint and the (f) minimum enclosed area constraints are computed by backpropagating through the rest of the mapping functions using the chain rule. The minimum area constraint seeks to eliminate violating regions, which corresponds to a negative gradient (red). The minimum hole constraint seeks to fill holes that are too small, which corresponds to a positive gradient (blue). All images were generated using resolutions that match the corresponding simulation resolution, which corresponds to small, pixelated artifacts in the design geometry, field patterns, and constraint gradients.

4. Minimum area and enclosed area constraints

We now define new minimum-area and minimum-enclosed-area constraints along with their respective gradients. The basic procedure for enforcing each design rule follows the methodology of the geometric constraints rather closely. First, indicator functions identify regions that violate the minimum-area constraint (“islands”) and regions that violate the minimum-enclosed-area constraint (“holes”). Next, each constraint will encourage the optimizer to either eliminate the violating regions via a localized erosion (for islands) or a localized dilation (for holes). Fig. 4 illustrates the two constraint functions on a sample design with their corresponding gradients.

We define the new minimum-area constraint function, g_A , as

$$g_A = \int \bar{\rho} I_A(\bar{\rho}) d\bar{\rho}, \quad (21)$$

where $\bar{\rho}$ are the projected design parameters and $I_A(\bar{\rho})$ is an indicator function that is nonzero (its “support”) in regions enclosing all the *islands* that are too small to be reliably fabricated. The process for determining these regions is described later. Similarly, we define the new minimum-enclosed-area constraint function, g_{EA} , as

$$g_{EA} = \int (1 - \bar{\rho}) I_{EA}(1 - \bar{\rho}) d\bar{\rho} \quad (22)$$

where I_{EA} is an indicator function that marks all the regions of the topology containing *holes* that are too small to be reliably fabricated.

To determine which regions (if any) violate the area and enclosed area rules, we use the marching-squares algorithm [38], a versatile image-segmentation technique often used to extract contours from images. The algorithm relies on a user-defined threshold point (between 0 and 1) such that “clusters” of pixels above the threshold point form part of the contour. In many ways, the procedure follows the same logic as the filter-threshold mapping used to evolve the design’s parameter field during optimization. For simplicity, we used an out-of-the-box implementation in Python [39] and a threshold parameter of 0.6.

Once the contours are identified, we calculate the area of each contour using a discrete summation of all the density values inside the contour, identified using morphological dilations. We used another out-of-the-box implementation in Python [40] to quickly and accurately compute each region’s area. If a particular contour has an area smaller than the prescribed design rule, the filled contour region is dilated by 1 pixel (e.g. using Eq. (9)) and subsequently added to the indicator function. As explained later, the minimal dilation conditions the gradient computation.

The constraints are then defined by

$$g_A \leq 0 \quad (23)$$

$$g_{EA} \leq 0 \quad (24)$$

such that the optimizer seeks to drive the constraint to zero. By doing so, the constraint *eliminates* regions that violate the area design rules, which means that violating islands must disappear (erode completely) and violating holes must be filled (dilate completely). Although one could alternatively imagine a more flexible approach that could *either* expand or contract the holes/islands, our purely elimination-based constraint greatly simplifies the resulting algorithm and its corresponding gradients. We note that the indicator functions and constraint functions will continuously decay to zero as the optimizer removes violating regions, a requirement for most gradient-based optimization algorithms. Since the indicator functions will reliably evaluate to zero if no violations are found, the above constraints can also be optimized to zero, a notable difference from the previously-defined geometric constraints that require a heuristically-determined convergence tolerance due to discretization error in the spatial gradient step [13].

The constraint gradients w.r.t. the filtered design parameters are calculated as

$$\frac{dg_A}{d\bar{\rho}} = I_A + \bar{\rho} \cdot \frac{dI_A}{d\bar{\rho}}, \quad (25)$$

$$\frac{dg_{EA}}{d\bar{\rho}} = -I_{EA} - (1 - \bar{\rho}) \cdot \frac{dI_{EA}}{d\bar{\rho}}. \quad (26)$$

To calculate the gradients w.r.t. the actual design parameters, we use the chain rule and backpropagate [30] through the thresholding and filtering steps

$$\frac{dg_A}{d\rho} = \frac{dg_A}{d\bar{\rho}} \frac{d\bar{\rho}}{d\tilde{\rho}} \frac{d\tilde{\rho}}{d\rho}, \quad (27)$$

$$\frac{dg_{EA}}{d\rho} = \frac{dg_{EA}}{d\bar{\rho}} \frac{d\bar{\rho}}{d\tilde{\rho}} \frac{d\tilde{\rho}}{d\rho}. \quad (28)$$

At first glance, it may seem that the indicator functions I_A and I_{EA} pose a severe challenge, because they are complicated to compute and are not always differentiable. However, it turns out that this difficulty is solved by the geometric constraints from Sec. 3, which ensure that the “islands” are well separated. In consequence, we can choose the boundaries of the indicator function’s support so that its boundaries occur only where $\bar{\rho}$ (for I_A) or $1 - \bar{\rho}$ (for I_{EA}) is

exponentially small, and hence changes in the indicator function make a negligible contribution to the constraint gradients. By slightly dilating each violating contour within the indicator function, we avoid any singularities associated with the edges of each contour. We can therefore safely omit this term from our analysis and still produce sufficiently accurate gradients. The gradients used for all numerical experiments in Sec. 6 are calculated using

$$\frac{dg_A}{d\rho} = I_A(\bar{\rho}) \frac{d\bar{\rho}}{d\rho} \frac{d\tilde{\rho}}{d\rho} \quad (29)$$

$$\frac{dg_{EA}}{d\rho} = -I_{EA}(1 - \bar{\rho}) \frac{d\bar{\rho}}{d\rho} \frac{d\tilde{\rho}}{d\rho} \quad (30)$$

Occasionally, the minimum-lengthscale constraints and the minimum-area constraints compete, effectively stalling the optimization. We overcome this process by weighting the two sets of constraints within the minimax optimization problem, as further explained in Sec. 6. This new formulation allows the optimizer to move freely between feasible regions without the usual issues that accompany stiff optimization (where the corresponding Hessian is ill-conditioned).

5. Robust optimization

In this section, we present a comprehensive approach to robust optimization that includes geometric constraints, area constraints, and an explicit under/over-etch tolerance. Commercial foundry lithography introduces yield variability, both random and systematic, throughout each step of the fabrication process. Current density-based robust optimization methodologies account for over/under-etch variability by running a worst-case optimization across three separate design fields [14, 25, 34]. The eroded and dilated design fields (which are geometric representations of over and under etching, respectively) are typically determined by choosing threshold parameters (η) that implicitly enforce quasi-lengthscales [14]. This approach can be problematic because all three design fields (eroded, dilated, and the nominal “blueprint”) must have the same topology [34], a strong requirement that may not hold in TO with small features. While some double-filtering techniques have shown potential for alleviating this restriction [15], predicting the actual amount of over- and under-etching relative to the desired lengthscale constraints is also difficult as it strongly depends on the dynamic range of the latent design parameters (ρ). In short, a reliable and deterministic method that decouples the lengthscale constraints from the prescribed etch tolerance has not yet been demonstrated.

We propose a straightforward approach that obeys arbitrary lengthscale and area design constraints, without any restrictions on the amount of over/under etching or on the topology itself. First, we place geometric constraints and minimum-area constraints on *just* the blueprint design. To ensure a predictable over/under-etch perturbation, we use the morphological transforms described in Sec. 2, such that the final erosions and dilations do not require any hyperparameter tuning (e.g. η) as they no longer depend on the dynamic range of the latent design field (illustrated in Fig. 3). Furthermore, this technique allows the topology to *change* between design fields without compromising the minimum length scales or requiring intermediary filter steps. The design rule constraints and robustness to manufacturing uncertainty are now completely decoupled, allowing the user to freely activate one or the other (although typically both will be applied).

Figure 5 compares a non-robust splitter to a robust splitter designed to be tolerant to roughly ± 20 nm of etch variation. Both splitters were optimized with the geometric constraints and area constraints described above. The robust variant requires two extra parallel Maxwell solves per device. As expected, the nominal design between the robust and non-robust variant show topological differences. The broadband performance (indicated in errorbars) for the robust splitter demonstrates greater tolerance than the non-robust variant even beyond the expected ± 20 nm variations.

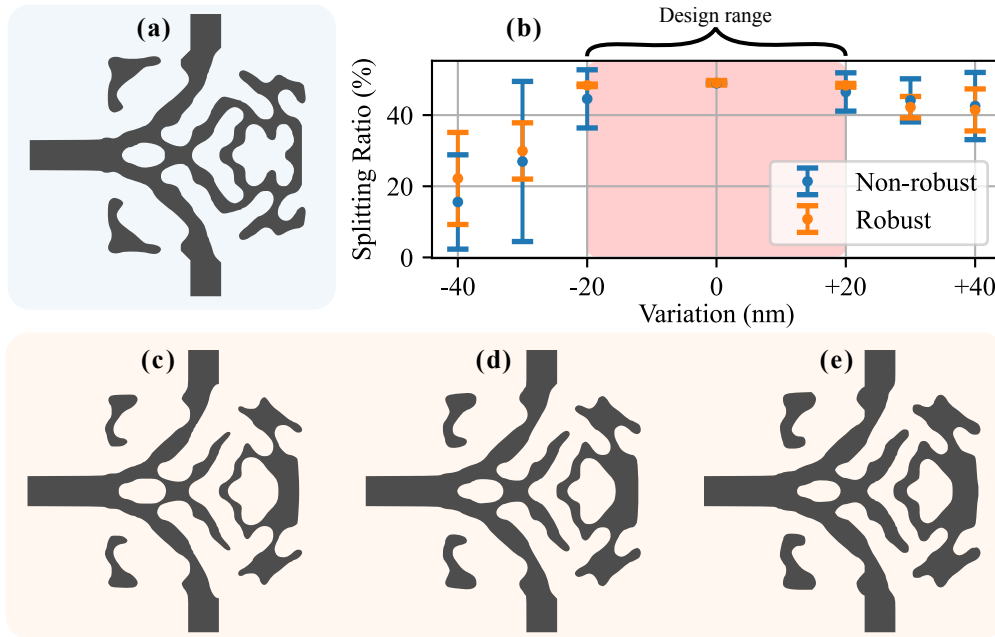


Fig. 5. Comparison of a robust T-splitter to a nominal T-splitter. (a) The nominal, non-robust design. (b) The broadband performance as a function of etching variation for both the robust and nominal design. The (c) eroded, (d) blueprint, (e) and dilated designs for the robust optimization. The final pixelated device geometries were smoothed (but not otherwise altered) to better illustrate the erosions and dilations

The devices were designed on a grid with 17 nm resolution. The resulting ± 20 nm target variation, along with the additional ± 30 nm and ± 40 nm perturbations, were approximated using the nonlinear harmonic filters described in Sec. 3.2. Like any spatial parameter, these values are only approximate once the grid discretization is included, but our 17 nm resolution is sufficient to clearly resolve distinct ± 20 nm, ± 30 nm, and ± 40 nm perturbations (filter radii of 1.2, 1.8, and 2.4 pixels), as shown in Fig. 5. Moreover, the point of performing robust optimization is that the results become less sensitive to small details such as the discretization resolution or the exact perturbation amount.

Note that the nominal design here actually has a certain degree of robustness simply by virtue of being broadband, which eliminates the extreme sensitivities that can arise in single-frequency optimization due to resonant and interference effects [41]. Therefore, it is not surprising that in Fig. 5 the nominal design still performs some splitting even for over/under-etching, and is qualitatively not too unlike the robust design. However, because the robust design is designed specifically to tolerate over/under-etching, its performance is quantitatively much better (its error bars are up to $24\times$ smaller for variations ≤ 20 nm) for that specific imperfection.

While this approach produces robust devices that simultaneously obey length scales, further work is needed to extend this methodology to *random* perturbations on the design field. Indeed existing methods that randomly vary the thresholding parameter η could be used, but they suffer from the same challenges discussed earlier.

6. Numerical examples

In this section, we present numerical results for three broadband silicon photonic examples to demonstrate the effectiveness and flexibility of our proposed methodology. Among the many photonic TO examples already discussed in the literature (1D grating couplers, demultiplexers, directional couplers, etc.) we chose (2D) devices that typically produce small islands and holes. Specifically, we demonstrate the design of several different waveguide reflectors, bends, and T-splitters because they each route light within a highly confined space ($3 \mu\text{m} \times 3 \mu\text{m}$). Each example uses the geometric and area constraints described earlier for nine different foundry specifications (for twenty-seven unique examples in total). While these specific design rules are synthetic, they are comparable to existing commercial-foundry DRC constraints.

Each example problem has an objective function depending on a waveguide modal overlap, which is proportional to an entry of the scattering matrix (S matrix):

$$\alpha_m^\pm = c \int_A [\mathbf{E}^*(r) \times \mathbf{H}_m^\pm(r) + \mathbf{E}_m^\pm(r) \times \mathbf{H}^*(r)] \cdot \hat{\mathbf{n}} dA \quad (31)$$

where α_m^\pm is the overlap coefficient (amplitude) of the m^{th} mode for the forward (+) and backward (-) directions, $\mathbf{E}(r)$ and $\mathbf{H}(r)$ are the Fourier-transformed total (simulated) fields at a particular frequency, $\mathbf{E}_m^\pm(r)$ and $\mathbf{H}_m^\pm(r)$ are the mode profiles at the same frequency for the forward- (+) and backward- (-) propagating modes, and c is a normalization constant chosen such that

$$|\alpha_m^\pm|^2 = P \quad (32)$$

where P is the total power propagating in that particular mode. Given these mode amplitudes $\alpha_m^\pm(\omega)$ at a frequency ω , each problem defines some objective function $f(\omega)$ to minimize, e.g. $f(\omega) = -|\alpha_1^+|^2$ to maximize the outgoing power in mode 1. For multiple frequencies ω_n , we will then minimize the maximum (worst-case) $f(\omega_n)$ using an epigraph formulation as explained in Sec. 2.

The fabrication constraints must be enforced “gradually” in the optimization problem, so that at early stages of the optimization the geometry can change freely (since changes in topology, such as the appearance of a new hole, temporarily violate the constraints [13]. For a fabrication constraint $g_k \leq 0$, this is accomplished by changing the constraint to $g_k \leq G_k$ for some constants $G_k \geq 0$ that are decreased as optimization proceeds:

$$\begin{aligned} & \min_{\rho} t \\ \text{s.t. } & \nabla \times \frac{1}{\mu_0 \mu_r} \nabla \times \mathbf{E} - \omega_m^2 \epsilon_0 \epsilon_r(\boldsymbol{\rho}) \mathbf{E} = -i\omega_m \mathbf{J} \quad n \in \{1, 2, \dots, 10\} \\ & 0 \leq \rho \leq 1 \quad (33) \\ & f(\omega_n) - t \leq 0 \quad n \in \{1, 2, \dots, 10\} \\ & g_k \leq G_k \quad k \in \{LS, LW, A, EA\} \end{aligned}$$

In many photonics-optimization problems, such as the examples here, the objective can be formulated so that $t \rightarrow \max_n f(\omega_n) \rightarrow 0$ as the optimization proceeds, and this can be used to simplify the choice of constraint bounds G_k . For example, if we are minimizing reflection or $1 - \text{transmission}$, then a high-performance structure will achieve near-zero optimum. In such cases, we can simply set $G_k = a_k t$ for a suitable constant a_k , so that $G_k \rightarrow 0$ as the optimization proceeds. Here, $a_k = 10^{-5}$ for $k \in \{LS, LW, A, EA\}$ worked for all the examples presented below. In the future, we are hopeful that such hyperparameters as G_k and the projection scale factor β can be automatically inferred from the convergence rate of the optimization algorithm. For now, however, they offer a strategic “knob” to turn when the optimization stalls or underperforms (further discussed in Sec. 6.3).

For the forward and adjoint solves, we used Meep, a free/open-source FDTD software package [29]. Each simulation ran with a resolution of 30 pixels/ μm ($\Delta x = 33$ nm) and a design-region (ρ) resolution of 60 pixels/ μm ($\Delta x = 17$ nm). Solid regions represent silicon with a dispersionless refractive index of $n = 3.4$. Void regions represent SiO_2 with a dispersionless refractive index of $n = 1.44$. We note that our method generalizes to materials with dispersion, and that dispersionless approximations were used here for simplicity. All examples were optimized using 70 iterations at $\beta = 8, 16,$ and 32 for a total of 210 iterations. The foundry constraints $g_k \leq G_k$ were not included until $\beta = 32$. We used a free/open-source implementation [42] of the globally convergent method of moving asymptotes (MMA) [35] optimization algorithm, restarting the algorithm each time β changed values. We ran each simulation on four cores of an Intel Xeon Gold 6226 2.7 GHz CPU using resources provided by the Partnership for an Advanced Computing Environment (PACE) at the Georgia Institute of Technology. Each example finished after about 6–8 hours of runtime.

6.1. Mirror

In this example, we designed a silicon photonic broadband mirror such that the forward-propagating mode is reflected, transferring energy to the backward-propagating mode of the same waveguide. Fig. 6 illustrates the design evolution of the broadband mirror, along with the steady state field pattern of the final device at the center of the band and the broadband reflection response. It is clear that a topology with several small holes and islands is favored by the optimization algorithm until the area and geometric constraints are activated during the last thresholding epoch (iteration 140, $\beta = 32$). At this point, the performance drops significantly and is gradually recovered while the foundry constraints are simultaneously satisfied. The final device exhibits *no* small holes or islands and obeys the prescribed linewidth, linespacing, and curvature rules, while still performing well across the band with $< 1\%$ variation in reflectivity.

The objective function used to design the mirror is

$$f(\omega_n) = 1 - |\alpha_1^-(\omega_n)|^2 \quad (34)$$

where α_1^- is the fundamental mode coefficient for the n^{th} frequency point (ω_n) propagating backwards. We perform a minimax optimization over the ten objective functions and the four foundry fabrication constraints. The fabrication constraint bounds $G_k = a_k t$ are weighted such that $a_k = 10^{-5}$.

Figure 7 illustrates the final designs: nine different broadband mirrors, each of which obey a different combination of fabrication constraints. The devices are presented such that those with the least-restrictive constraints are on the upper-left side of the figure and the constraints become more restrictive toward the bottom-right. The least-restrictive rulebook, for example, imposes a minimum linewidth/linespacing of 60 nm and a minimum area/enclosed area of $0.005 \mu\text{m}^2$, and produced a device shown in the upper-left corner. In contrast, the design shown in the lower-right corner obeys a minimum linewidth/linespacing of 120 nm and a minimum area/enclosed area of $0.2 \mu\text{m}^2$. The devices located between these two boundary cases are intermediate versions of these design rules, including cases where linewidth \neq linespacing and/or area \neq enclosed area.

All but two of the resulting devices reflect $> 90\%$ of the light with very little variation across the band (the most restrictive rulebook produced a device with just $\pm 0.3\%$ of variation across the band). Because these algorithms produce only a *local* optimum, not a global optimum, the devices that have less restrictive constraints (upper-left) *do not* always outperform the devices designed with tighter restrictions (lower-right). For example, the device with a linewidth/linespacing constraint of 90 nm and a minimum area/enclosed area constraint of $0.005 \mu\text{m}^2$ only achieve an average of 68.8% transmission with a rather high variability of 18.7% across the band. A global optimizer subject to this particular rulebook could have also produced the same design as the device on its right, as it shares the same linewidth/linespacing constraint (and thus the same filter

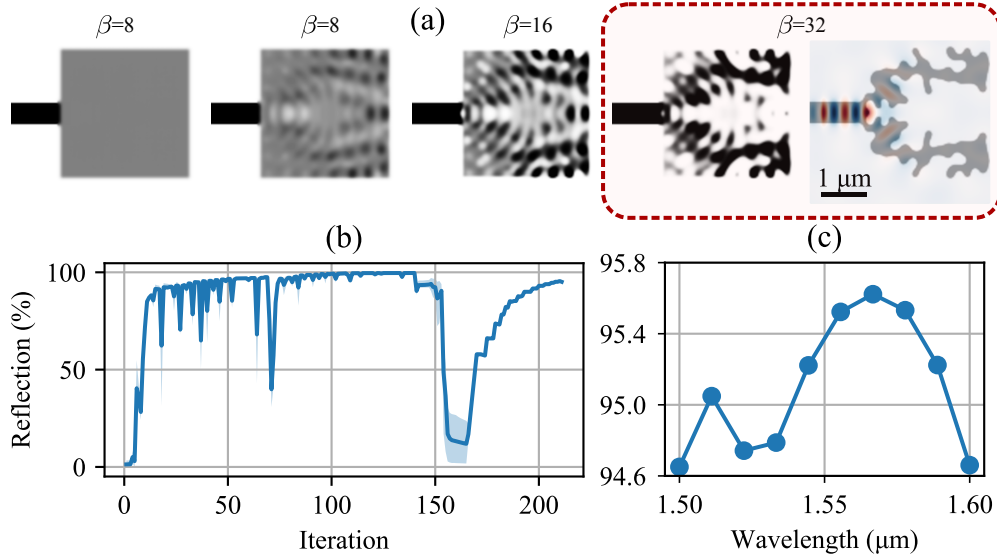


Fig. 6. Evolution and performance of the broadband mirror designed to obey a minimum linewidth of 90 nm, a minimum spacing of 90 nm, a minimum area of $0.08 \mu\text{m}^2$, and a minimum enclosed area of $0.2 \mu\text{m}^2$. (a) Projected design field ($\bar{\rho}$) evolution for iterations 1, 31, 81, 146, and 210. The final iteration (210) also displays the steady-state field pattern of the device at $\lambda=1.55 \mu\text{m}$. The two iterations enclosed in red indicate when the constraint was active at $\beta = 32$. (b) The evolution of the device broadband reflection as a function of iteration. The mean performance across the band is depicted by a solid line, while the bounds are illustrated by a blue shaded region. All constraints were enforced starting with iteration 141. (c) The broadband performance of the device from $1.50 \mu\text{m}$ to $1.60 \mu\text{m}$.

radius) but with slightly stricter area constraints (which still satisfy a less-strict constraint by definition). However, as seen here, a local optimizer may happen to find a poorer local optimum in a case with looser constraints, and it is well known that some experimentation (e.g. different starting points or different constraint weights) is often needed to find a satisfactory local optimum in high-dimensional nonconvex optimization (see Sec. 6.3). Nevertheless, as is often observed in topology optimization [20], in all cases the algorithm converged to a local optimum with good performance.

6.2. Bend

In the next example, we design a silicon photonic broadband 90° bend in a single-mode waveguide. Fig. 8 illustrates the design evolution of the broadband bend, along with the steady-state field pattern of the final device at the center of the band and the broadband transmission response. As expected, the design converges to a binary layout that satisfies the prescribed DRC rules. Unlike the mirror examples earlier, the optimization favored larger, elongated structures/gaps that satisfied the area rules even before the constraints were applied. As such, the drop in performance at the start of the final epoch is only about 10%.

The objective function used to design the bend is

$$f(\omega_n) = 1 - |\alpha_1^+|^2 \quad (35)$$

where α_1^+ is the fundamental mode coefficient for the n^{th} frequency point propagating forward (North of the bend).

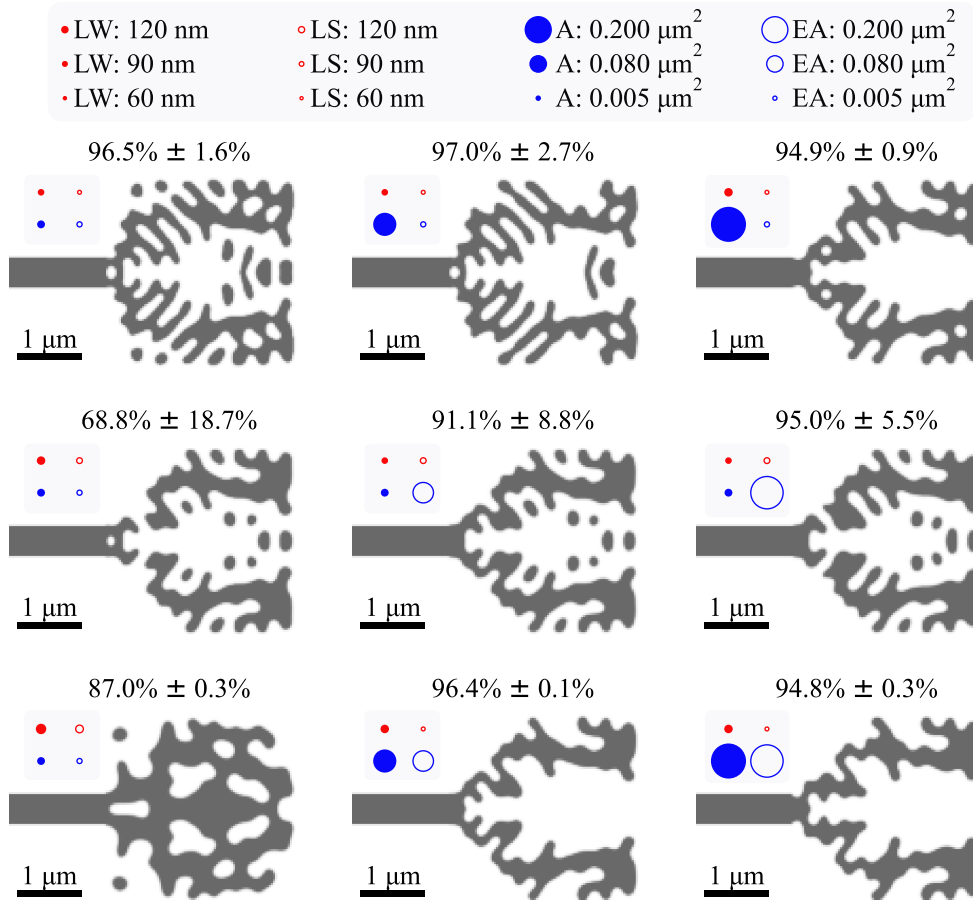


Fig. 7. Comparison of various broadband mirror designs optimized to perform under different combinations of fabrication constraints. (top) Each of the minimum linewidth, minimum spacing, minimum area, and minimum enclosed area constraints drawn to scale. The minimum area constraints, for example, are demonstrated by a circle that shares the same area as the constraint. The minimum linewidth and spacing constraints are demonstrated by circles whose diameters align with the corresponding constraint. By extension, the minimum radius of curvature also aligns with this depiction. (bottom) Each broadband mirror design is drawn alongside the constraints used to produce it. The average reflection of each device from $\lambda = 1.5 \mu\text{m}$ to $\lambda = 1.6 \mu\text{m}$, along with the corresponding minimum and maximum is depicted above each design.

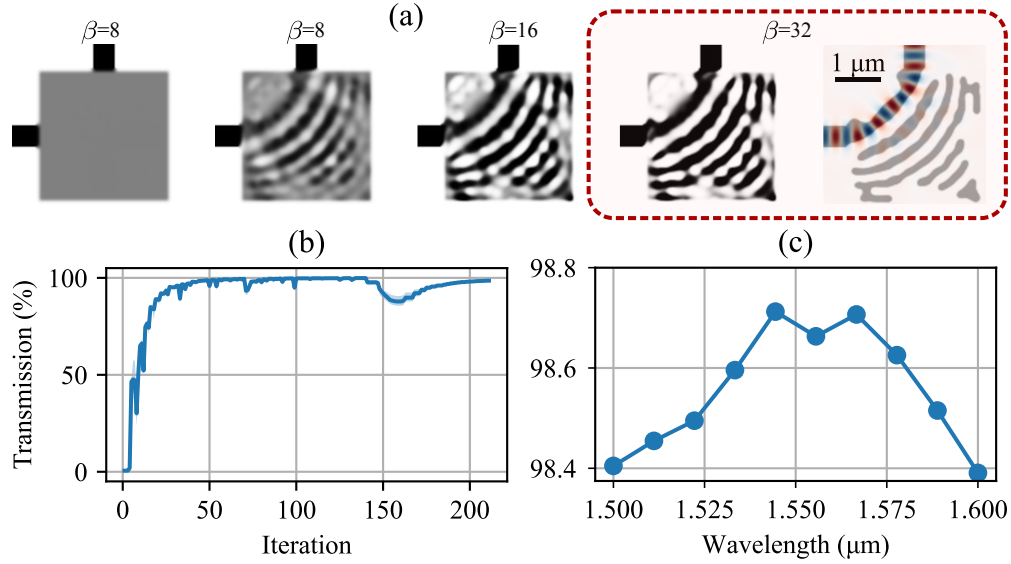


Fig. 8. Evolution and performance of the broadband bend designed to obey a minimum linewidth of 90 nm, a minimum spacing of 90 nm, a minimum area of $0.08 \mu\text{m}^2$, and a minimum enclosed area of $0.2 \mu\text{m}^2$. (a) Projected design field ($\bar{\rho}$) evolution for iterations 1, 31, 81, 146, and 210. The final iteration (210) also displays the steady-state field pattern of the device at $\lambda = 1.55 \mu\text{m}$. The two iterations enclosed in red indicate when the constraint was active at $\beta = 32$. (b) The evolution of the device broadband reflection as a function of iteration. The mean performance across the band is depicted by a solid line, while the bounds are illustrated by a blue shaded region. All constraints were enforced starting with iteration 141. (c) The broadband performance of the device is from $\lambda = 1.50 \mu\text{m}$ to $\lambda = 1.60 \mu\text{m}$.

Figure 9 shows the final designs for nine different broadband bends, each of which obeys a different combination of fabrication constraints (just as for the mirror designs above). Again, the devices perform well across the various rulebooks, with the average transmission always $> 89\%$. We note that the most constrained device (lower-right corner) exhibits a transmission of 98.2% with practically no variation across the band. Interestingly, the area constraints removed the small Bragg mirror-like features on the boundaries as soon as the constraints were activated. The resulting device naturally took a rather different optimization evolution than the other devices constrained by different rulebooks.

6.3. T-splitter

In the final example, we design a symmetric, silicon-photonic broadband T-splitter. Unlike conventional splitters that route the outputs along the same direction as they entered the splitting structure, the outputs of this design are each routed at a 90° angle (forming a "T"). Consequently, we intuitively expect structures containing small islands and holes in order to reflect the light within the device and route it as intended. Fig. 10 illustrates the design evolution of the broadband splitter, along with the steady-state field pattern of the final device at the center of the band and the broadband transmission response. Similar to the mirror example above, we first see several small islands forming in the initial two epochs ($\beta = 8, 16$). Once the constraints are applied, however, these violating regions are modified such that they no longer pose any DRC issues. Specifically, the geometric constraints "bridged" the regions that were within the rulebook's defined linespacing, while the area constraints eliminated the remaining well-separated islands.

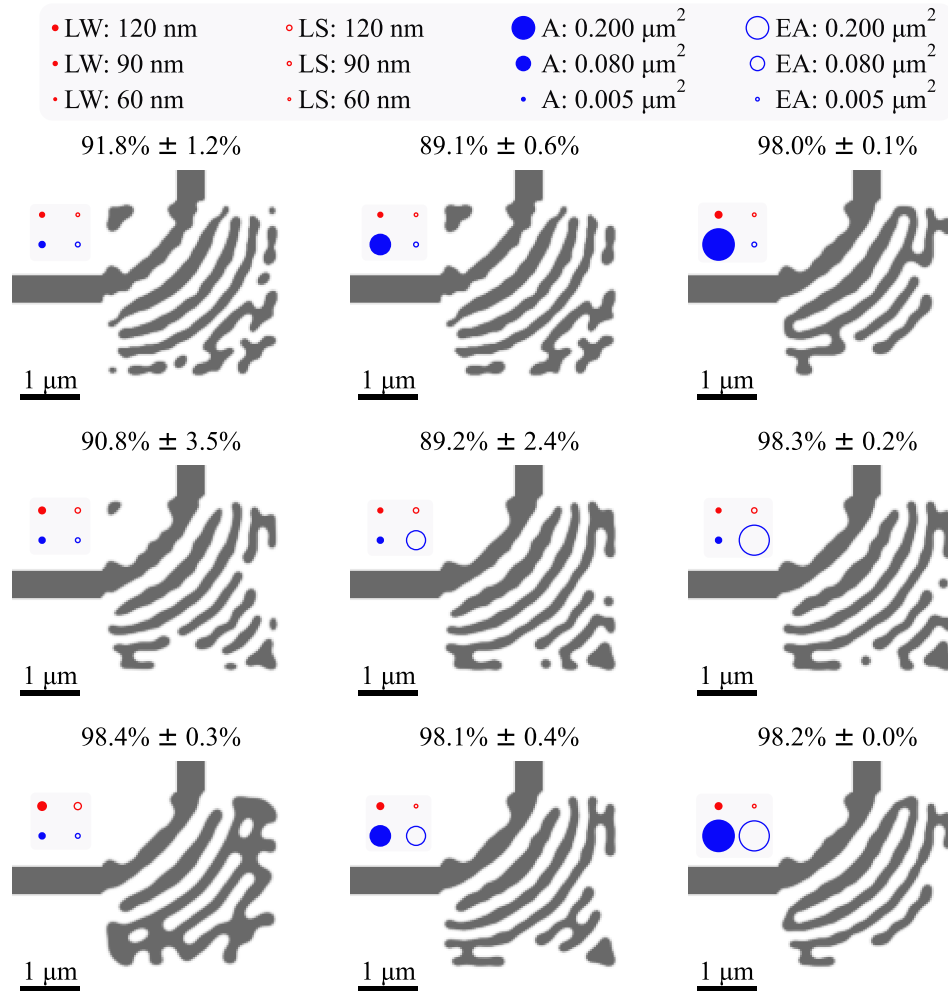


Fig. 9. Comparison of various broadband bend designs optimized to perform under different combinations of fabrication constraints. (top) Each of the minimum linewidth, minimum spacing, minimum area, and minimum enclosed area constraints drawn to scale. The minimum area constraints, for example, are demonstrated by a circle that shares the same area as the constraint. The minimum linewidth and spacing constraints are demonstrated by circles whose diameters align with the corresponding constraint. By extension, the minimum radius of curvature also aligns with this depiction. (bottom) Each broadband mirror design is drawn alongside the constraints used to produce it. The average transmission of each device from $\lambda = 1.5 \mu\text{m}$ to $\lambda = 1.6 \mu\text{m}$, along with the corresponding minimum and maximum is depicted above each design.

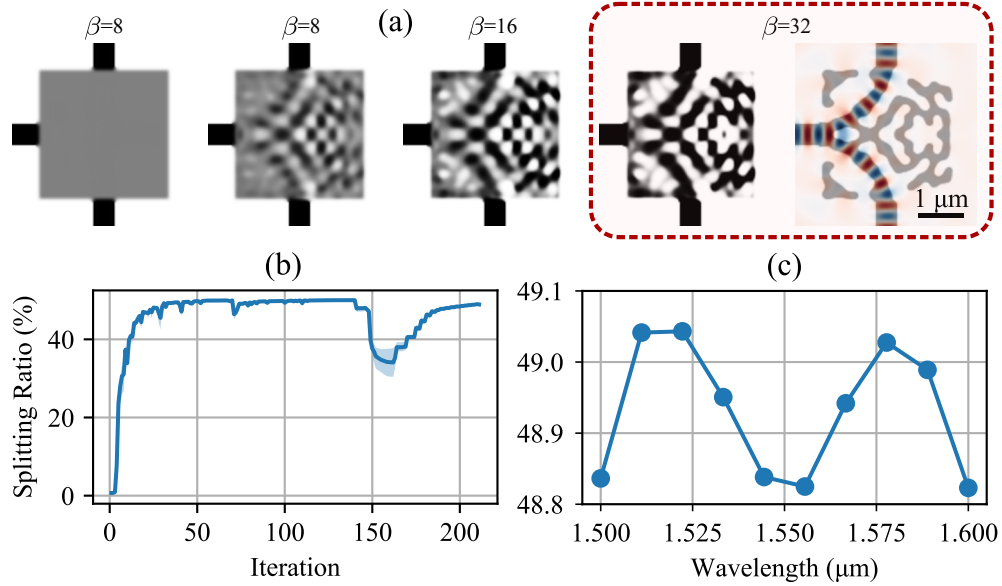


Fig. 10. Evolution and performance of the T-splitter designed to obey a minimum linewidth of 90 nm, minimum spacing of 90 nm, a minimum area of $0.08 \mu\text{m}^2$, and minimum enclosed area of $0.08 \mu\text{m}^2$. (a) Projected design field ($\bar{\rho}$) evolution for iterations 1, 30, 80, 145, and 210. The two iterations enclosed in red indicate when the constraint was active at $\beta = 32$. (b) The evolution of the device splitting ratio as a function of iteration. All constraints were enforced starting with iteration 140. (c) The steady state field pattern of the final device at $\lambda = 1.55 \mu\text{m}$. (d) The broadband performance of the device from $\lambda = 1.50 \mu\text{m}$ to $\lambda = 1.60 \mu\text{m}$.

The enclosed-area constraints ensured that no new small holes would form as a result of this bridging process. The resulting device exhibits 0.2% of variation across the band with a mean transmission of 49%.

The objective function used to design the T-splitter is

$$f(\omega_n) = 1 - |\alpha_{(1,1)}^+|^2 - |\alpha_{(1,2)}^+|^2. \quad (36)$$

We imposed a geometric symmetry constraint, forcing the final splitting ratio to *also* be symmetric.

Figure 11 illustrates the final designs for nine different broadband splitters, each of which obeys a different combination of fabrication constraints (just as for the mirror and bend designs above). In this case, we see that the least-restricted device (upper-left corner) happens to exhibit rather poor performance across the band ($27.2\% \pm 5.7\%$), which turns out to simply be an unlucky local optimum. We reran the exact same optimization problem for this rulebook, but modified the constraint weight factors to $a_k = 10^{-3}$, effectively relaxing the strength of each constraint (without modifying the constraint function itself). The resulting device still obeyed the prescribed design rule constraints, but exhibited far superior performance across the band ($49.5\% \pm 0.05\%$). In general, we found that if the constraint weights were too strict, the optimizer could become stuck in a local minimum. The constraint weights (a_k) offer additional tunability to combat slow convergence.

7. Conclusion

We presented a comprehensive photonic TO design methodology that includes minimum linewidth, linespacing, curvature, area, and enclosed-area “DRC” constraints. The linewidth,

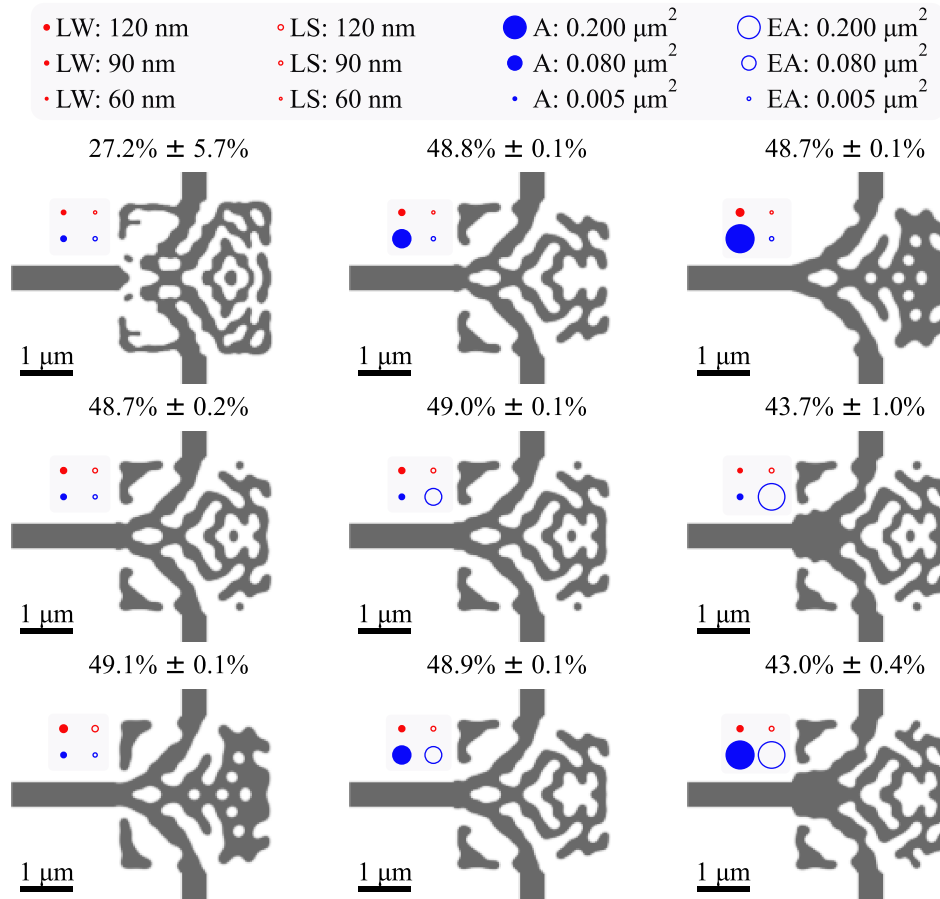


Fig. 11. Comparison of various broadband T-splitter designs optimized to perform under different combinations of fabrication constraints. (top) Each of the minimum linewidth, minimum spacing, minimum area, and minimum enclosed area constraints drawn to scale. The minimum area constraints, for example, are demonstrated by a circle that shares the same area as the constraint. The minimum linewidth and spacing constraints are demonstrated by circles whose diameters align with the corresponding constraint. By extension, the minimum radius of curvature also aligns with this depiction. (bottom) Each broadband mirror design is drawn alongside the constraints used to produce it. The average transmission of each device from $\lambda = 1.5 \mu\text{m}$ to $\lambda = 1.6 \mu\text{m}$, along with the corresponding minimum and maximum is depicted above each design.

linespacing, and curvature constraints are adapted from previous work and combined with novel area constraints. All the constraints rely on density-based projection steps without any additional re-parameterizations. By additionally incorporating differentiable morphological transforms, we provide a method to perform systematic robust optimization while simultaneously imposing the fabrication constraints. We illustrated these techniques using three distinct test devices constrained by nine different design rulebooks.

There remain many opportunities for future work on these topics. As we noted in Sec. 4, it is desirable to have a more flexible area constraint that allows the topology-violating region to either expand or contract, instead of only contracting as in this work. Similarly, a more rigorous analysis that relates the hyperparameters of the geometric constraints (G_k) is needed, such that the need to tune each optimization run could be reduced. One possible path forward involves deriving an improved constraint function based on the morphological transforms described in Sec. 3.2. While our new robust-optimization paradigm decouples systematic under/over-etch variations from DRC constraints, future efforts should seek to handle a larger class of manufacturing uncertainties, such as over-etch in some regions and under-etch in others, and surface roughness, within a similar morphological-transform framework. Similarly, our methodology does not explicitly incorporate calibration steps that are typically performed by the foundry, such as optical proximity correction (OPC). While recent work has demonstrated a variant of TO that uses filters to mitigate the need for OPC [43, 44], most semiconductor foundry pipelines (and their respective DRC rulebooks) are not set up for this workflow.

Funding. AMH was supported by the Department of Defense (DoD) through the National Defense Science & Engineering Graduate (NDSEG) Fellowship Program. AO was partially supported by the National Science Foundation (NSF) via Small Business Innovation Research (SBIR) Phase I and II awards 1647206 and 1758596. SGJ was supported in part by the U.S. Army Research Office through the Institute for Soldier Nanotechnologies (Award No. W911NF-18-2-0048). SER was supported by the Georgia Electronic Design Center of the Georgia Institute of Technology.

Acknowledgments. The authors would like to thank Rasmus E. Christiansen for his discussions while completing this work. This research was supported in part through research cyberinfrastructure resources and services provided by the Partnership for an Advanced Computing Environment (PACE) at the Georgia Institute of Technology, Atlanta, Georgia, USA.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. A. Y. Piggott, J. Petykiewicz, L. Su, and J. Vučković, “Fabrication-constrained nanophotonic inverse design,” *Sci. reports* **7**, 1–7 (2017).
2. L. Su, R. Trivedi, N. V. Sapra, A. Y. Piggott, D. Vercruysse, and J. Vučković, “Fully-automated optimization of grating couplers,” *Opt. express* **26**, 4023–4034 (2018).
3. N. V. Sapra, D. Vercruysse, L. Su, K. Y. Yang, J. Skarda, A. Y. Piggott, and J. Vučković, “Inverse design and demonstration of broadband grating couplers,” *IEEE J. Sel. Top. Quantum Electron.* **25**, 1–7 (2019).
4. C. Dory, D. Vercruysse, K. Y. Yang, N. V. Sapra, A. E. Rugar, S. Sun, D. M. Lukin, A. Y. Piggott, J. L. Zhang, M. Radulaski, K. G. Lagoudakis, L. Su, and J. Vučković, “Inverse-designed diamond photonics,” *Nat. communications* **10**, 1–7 (2019).
5. F. Wang, R. E. Christiansen, Y. Yu, J. Mørk, and O. Sigmund, “Maximizing the quality factor to mode volume ratio for ultra-small photonic crystal cavities,” *Appl. Phys. Lett.* **113**, 241101 (2018).
6. A. Y. Piggott, J. Lu, K. G. Lagoudakis, J. Petykiewicz, T. M. Babinec, and J. Vučković, “Inverse design and demonstration of a compact and broadband on-chip wavelength demultiplexer,” *Nat. Photonics* **9**, 374–377 (2015).
7. L. Su, A. Y. Piggott, N. V. Sapra, J. Petykiewicz, and J. Vučković, “Inverse design and demonstration of a compact on-chip narrowband three-channel wavelength demultiplexer,” *Acs Photonics* **5**, 301–305 (2018).
8. K. Y. Yang, J. Skarda, M. Cotrufo, A. Dutt, G. H. Ahn, M. Sawaby, D. Vercruysse, A. Arbabian, S. Fan, A. Alù, and J. Vučković, “Inverse-designed non-reciprocal pulse router for chip-based lidar,” *Nat. Photonics* **14**, 369–374 (2020).
9. T. W. Hughes, M. Minkov, I. A. D. Williamson, and S. Fan, “Adjoint method and inverse design for nonlinear nanophotonic devices,” *ACS Photonics* **5**, 4781–4787 (2018).

10. R. E. Christiansen, Z. Lin, C. Roques-Carmes, Y. Salamin, S. E. Kooi, J. D. Joannopoulos, M. Soljačić, and S. G. Johnson, "Fullwave maxwell inverse design of axisymmetric, tunable, and multi-scale multi-wavelength metalenses," *Opt. Express* **28**, 33854–33868 (2020).
11. R. E. Christiansen, J. Michon, M. Benzaouia, O. Sigmund, and S. G. Johnson, "Inverse design of nanoparticles for enhanced raman scattering," *Opt. express* **28**, 4444–4462 (2020).
12. D. Vercauteren, N. V. Sapra, L. Su, R. Trivedi, and J. Vučković, "Analytical level set fabrication constraints for inverse design," *Sci. reports* **9**, 1–7 (2019).
13. M. Zhou, B. S. Lazarov, F. Wang, and O. Sigmund, "Minimum length scale in topology optimization by geometric constraints," *Comput. Methods Appl. Mech. Eng.* **293**, 266–282 (2015).
14. F. Wang, J. S. Jensen, and O. Sigmund, "Robust topology optimization of photonic crystal waveguides with tailored dispersion properties," *JOSA B* **28**, 387–397 (2011).
15. R. E. Christiansen, B. S. Lazarov, J. S. Jensen, and O. Sigmund, "Creating geometrically robust designs for highly sensitive problems using topology optimization," *Struct. Multidiscip. Optim.* **52**, 737–754 (2015).
16. L. Chrostowski and M. Hochberg, *Silicon photonics design: from devices to systems* (Cambridge University, 2015).
17. W. Bogaerts and L. Chrostowski, "Silicon photonics circuit design: methods, tools and challenges," *Laser & Photonics Rev.* **12**, 1700237 (2018).
18. L. Hägg and E. Wadbro, "On minimum length scale control in density based topology optimization," *Struct. Multidiscip. Optim.* **58**, 1015–1032 (2018).
19. M. Liehr, M. Baier, G. Hoefler, N. M. Fahrenkopf, J. Bowers, R. Gladhill, P. O'Brien, E. Timurdogan, Z. Su, and F. Kish, "Foundry capabilities for photonic integrated circuits," in *Optical Fiber Telecommunications VII*, (Elsevier, 2020), pp. 143–193.
20. O. Sigmund and K. Maute, "Topology optimization approaches," *Struct. Multidiscip. Optim.* **48**, 1031–1055 (2013).
21. E. Khoram, X. Qian, M. Yuan, and Z. Yu, "Controlling the minimal feature sizes in adjoint optimization of nanophotonic devices using b-spline surfaces," *Opt. express* **28**, 7060–7069 (2020).
22. X. Qian, "Topology optimization in b-spline space," *Comput. Methods Appl. Mech. Eng.* **265**, 15–35 (2013).
23. Y. M. Yoely, I. Hanniel, and O. Amir, "Structural optimization with explicit geometric constraints using a b-spline representation," *Mech. Based Des. Struct. Mach.* pp. 1–32 (2020).
24. M. Chen, J. Jiang, and J. A. Fan, "Design space reparameterization enforces hard geometric constraints in inverse-designed nanophotonic devices," *ACS Photonics* **7**, 3141–3151 (2020).
25. E. W. Wang, D. Sell, T. Phan, and J. A. Fan, "Robust design of topology-optimized metasurfaces," *Opt. Mater. Express* **9**, 469–482 (2019).
26. X. Qian and O. Sigmund, "Topological design of electromechanical actuators with robustness toward over-and under-etching," *Comput. Methods Appl. Mech. Eng.* **253**, 237–251 (2013).
27. C. M. Lalau-Keraly, S. Bhargava, O. D. Miller, and E. Yablonovitch, "Adjoint shape optimization applied to electromagnetic design," *Opt. express* **21**, 21693–21701 (2013).
28. O. D. Miller, "Photonic design: From fundamental solar cell physics to computational inverse design," arXiv preprint arXiv:1308.0212 (2013).
29. A. F. Oskooi, D. Roundy, M. Ibanescu, P. Bermel, J. D. Joannopoulos, and S. G. Johnson, "Meep: A flexible free-software package for electromagnetic simulations by the ftdtd method," *Comput. Phys. Commun.* **181**, 687–702 (2010).
30. B. S. Lazarov, F. Wang, and O. Sigmund, "Length scale and manufacturability in density-based topology optimization," *Arch. Appl. Mech.* **86**, 189–218 (2016).
31. Y. Cao, S. Li, L. Petzold, and R. Serban, "Adjoint sensitivity analysis for differential-algebraic equations: The adjoint dae system and its numerical solution," *SIAM journal on scientific computing* **24**, 1076–1089 (2003).
32. L. Su, D. Vercauteren, J. Skarda, N. V. Sapra, J. A. Petykiewicz, and J. Vučković, "Nanophotonic inverse design with spins: Software architecture and practical considerations," *Appl. Phys. Rev.* **7**, 011407 (2020).
33. L. Alloatti, M. Wade, V. Stojanovic, M. Popovic, and R. J. Ram, "Photonics design tool for advanced cmos nodes," *IET optoelectronics* **9**, 163–167 (2015).
34. M. Schevenels, B. S. Lazarov, and O. Sigmund, "Robust topology optimization accounting for spatially varying manufacturing errors," *Comput. Methods Appl. Mech. Eng.* **200**, 3613–3627 (2011).
35. K. Svanberg, "A class of globally convergent optimization methods based on conservative convex separable approximations," *SIAM journal on optimization* **12**, 555–573 (2002).
36. 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).
37. K. Svanberg and H. Svärd, "Density filters for topology optimization based on the pythagorean means," *Struct. Multidiscip. Optim.* **48**, 859–875 (2013).
38. W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3d surface construction algorithm," *ACM siggraph computer graphics* **21**, 163–169 (1987).
39. S. Van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, and T. Yu, "scikit-image: image processing in python," *PeerJ* **2**, e453 (2014).
40. P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Í. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman,

- I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nat. Methods* **17**, 261–272 (2020).
41. A. Mutapcic, S. Boyd, A. Farjadpour, S. G. Johnson, and Y. Avniel, "Robust design of slow-light tapers in periodic waveguides," *Eng. Optim.* **41**, 365–384 (2009).
 42. S. G. Johnson, "The NLOpt nonlinear-optimization package," <http://github.com/stevengj/nlopt>.
 43. M. Zhou, B. S. Lazarov, and O. Sigmund, "Topology optimization for optical projection lithography with manufacturing uncertainties," *Appl. optics* **53**, 2720–2729 (2014).
 44. M. Zhou, B. S. Lazarov, and O. Sigmund, "Topology optimization for optical microlithography with partially coherent illumination," *Int. J. for Numer. Methods Eng.* **109**, 631–647 (2017).