Ultrafast Simulation and Optimization of Nanophotonic Devices with Integral Equation Methods: Supporting Information

Constantine Sideris,*,† Emmanuel Garza,‡ and Oscar P. Bruno‡

†*Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA 90089, USA*

‡*Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA 91125*

E-mail: csideris@usc.edu

Contents

SI Incident excitations

In the context of silicon photonic device simulation Gaussian beam excitations are often used as approximations of the field output from the fundamental mode of a fiber [S1]. In a Gaussian beam the incident excitation $u^{\text{inc}}_{\text{beam}}$ is a solution to the paraxial approximation of the Helmholtz equation [S2], and therefore the corresponding field $(\boldsymbol{E}^{\text{inc}}_{\text{beam}}, \boldsymbol{H}^{\text{inc}}_{\text{beam}})$ does not exactly satisfy Maxwell's equations. However, if defined properly and across short enough propagation distances, the scalar Gaussian field, which is given by

$$u^{\text{inc}}_{\text{beam}}(z, x) = \exp\left(-\frac{z^2}{w_0^2}\right) \tag{S1}$$

(where w_0 denotes the Gaussian beam half-width), does provide an adequately accurate approximation for the incident field arising from an optical fiber.

The incident excitation fields u^{inc} on a given SIW, on the other hand, can be selected as an arbitrary linear combination

$$u^{\text{inc}}(z, x) = \sum_{m=1}^{M} A^{\text{inc}}_m u^m_\perp(x) e^{\mathrm{i}k^m_z z} \tag{S2}$$

of the finitely-many modes supported by the waveguide. Here A^{inc}_m is a complex number, $u^m_\perp$ is the transverse profile of the m-th mode of the waveguide, and k^m_z is the corresponding propagation constant. As discussed in the next section regarding waveguide mode solutions, $u^m_\perp$ and k^m_z often have analytical or semi-analytical forms, and they can otherwise always be

obtained by numerical solution of simple one dimensional eigenvalue problems [S3].

As indicated in [S4], calling Ω^{inc} the union of all SIW regions which are used to convey a non-vanishing incident field, letting $\boldsymbol{r} = (z, x)$, and defining χ^{inc} as the indicator function of the set Ω^{inc},

$$\chi^{\text{inc}}(\boldsymbol{r}) = \begin{cases} 1 & \text{for } \boldsymbol{r} \in \Omega^{\text{inc}} \\ 0 & \text{for } \boldsymbol{r} \notin \Omega^{\text{inc}}, \end{cases} \tag{S3}$$

the total electric field $\boldsymbol{E}$ can be expressed in the form $\boldsymbol{E} = \boldsymbol{E}^{\text{inc}}\chi^{\text{inc}} + \boldsymbol{E}^{\text{scat}}$, where the scattered field $\boldsymbol{E}^{\text{scat}}$ is an outgoing (radiating) solution of Maxwell's equations [S4, 5]. Throughout this paper the temporal dependence is assumed to be time-harmonic, and the corresponding exponential factors $e^{-i\omega t}$ are assumed throughout, but otherwise omitted.

As mentioned in the Boundary Integral Formulation subsection in the Results of the main text, the densities ϕ and ψ, used in the representation formula (eq. 5, main text), can be expressed as a sum of their incident and scattered components: $\phi = \phi^{\text{inc}} + \phi^{\text{scat}}$ and $\psi = \psi^{\text{inc}} + \psi^{\text{scat}}$ where, using the indicator function (S3), we obtain

$$\begin{aligned} \phi^{\text{inc}}(\boldsymbol{r}) \equiv \chi^{\text{inc}}(\boldsymbol{r}) u_+^{\text{inc}}(\boldsymbol{r}) \quad \text{and} \quad \psi^{\text{inc}}(\boldsymbol{r}) \equiv \chi^{\text{inc}} \frac{\partial u_+^{\text{inc}}(\boldsymbol{r})}{\partial \boldsymbol{n}}, \\ \phi^{\text{scat}}(\boldsymbol{r}) \equiv u_+^{\text{scat}}(\boldsymbol{r}) \quad \text{and} \quad \psi^{\text{scat}}(\boldsymbol{r}) \equiv \frac{\partial u_+^{\text{scat}}(\boldsymbol{r})}{\partial \boldsymbol{n}}. \end{aligned} \quad \text{for } \boldsymbol{r} \in \Gamma. \tag{S4}$$

SII Waveguide modes

Some of the examples considered in this work rely on the modes of the input and output waveguides guiding light into and out of the photonic devices, such as those depicted in Fig. 1 of the main text. These waveguides contains a core guiding layer of index n_{co} and wavenumber k_{co} surrounded by cladding material of dielectric constant n_{cl} and wavenumber

$k_{\rm cl}$. The transverse modal profile for such a waveguide is given by

$$u_{\perp}^{m}(x) = \begin{cases} \nu(\gamma_{\rm co}h)e^{-\gamma_{\rm cl}(x-h)} & (x > h) \\ \nu(\gamma_{\rm co}x) & (|x| \leq h) \\ \nu(-\gamma_{\rm co}h)e^{\gamma_{\rm cl}(x+h)} & x < -h. \end{cases} \tag{S5}$$

Here we have assumed that the guide has half-width h and is centered at $x = 0$, we have set $\gamma_{\rm co} = \sqrt{k_{\rm co}^2 - (k_z^m)^2}$, $\gamma_{\rm cl} = \sqrt{(k_z^m)^2 - k_{\rm cl}^2}$, $\nu(\tau) = \cos(\tau)$ or $\nu(\tau) = \sin(\tau)$) (for symmetric modes and antisymmetric modes, respectively), and, letting $\nu_{\rm wg} = 1$ for TE polarization and $\nu_{\rm wg} = (\frac{k_{\rm cl}}{k_{\rm co}})^2$ for TM polarization, we have denoted by k_z^m the m-th solution of the transcendental equation

$$\begin{cases} \gamma_{\rm cl} = \nu_{\rm wg}\gamma_{\rm co}\tan(\gamma_{\rm co}h) & \text{(symmetric)} \\ \gamma_{\rm cl} = -\nu_{\rm wg}\gamma_{\rm co}\text{cotan}(\gamma_{\rm co}h) & \text{(antisymmetric).} \end{cases} \tag{S6}$$

SIII Mode Overlap Integrals

The design of silicon photonic devices often requires evaluation of the power guided by input and output waveguides. Calling the positive and negative x directions the "forward" and "backward" propagation directions, respectively, the electromagnetic field propagating along an infinite waveguide is given by [S7]

$$\begin{aligned} \boldsymbol{E} &= \sum_m \left(a_m \boldsymbol{E}_m^{\rm fwd} + b_m \boldsymbol{E}_m^{\rm bwd}\right) + \boldsymbol{E}^{\rm rad} \\ &= \sum_m \left(a_m \boldsymbol{E}_m + b_m \overline{\boldsymbol{E}}_m\right) + \boldsymbol{E}^{\rm rad} \\ \boldsymbol{H} &= \sum_m \left(a_m \boldsymbol{H}_m^{\rm fwd} + b_m \boldsymbol{H}_m^{\rm bwd}\right) + \boldsymbol{H}^{\rm rad} \\ &= \sum_m \left(a_m \boldsymbol{H}_m - b_m \overline{\boldsymbol{H}}_m\right) + \boldsymbol{H}^{\rm rad} \end{aligned} \tag{S7}$$

where, letting $(\boldsymbol{E}_m^{\text{fwd}}, \boldsymbol{H}_m^{\text{fwd}}) = (\boldsymbol{E}_m, \boldsymbol{H}_m)$ denote the m-th forward-propagating electromagnetic waveguide mode, where, using overbars to denote conjugates, $(\boldsymbol{E}_m^{\text{bwd}}, \boldsymbol{H}_m^{\text{bwd}}) = (\overline{\boldsymbol{E}}_m, -\overline{\boldsymbol{H}}_m)$ is the m-th backward propagating mode, and where a_m and b_m denote the forward and backward amplitude coefficients, respectively. The radiating field $(\boldsymbol{E}^{\text{rad}}, \boldsymbol{H}^{\text{rad}})$, which are part of the spectral decomposition of a general electromagnetic field along dielectric waveguide, decay along the waveguide—so that only the forward and backward modes need to be considered sufficiently far from any inhomogeneities in large but finite waveguide. All mode overlap calculations used below in this paper are performed at points far along the waveguides from any waveguide inhomogeneities, so that it can be assumed that the radiating fields (i.e., fields additional to all existing propagating modes) are negligible. The mode profiles for a loss-less infinite waveguide are orthogonal to each other and they are assumed to be normalized in such a way that the relation

$$\int_{-\infty}^{+\infty} \left(\boldsymbol{E}_m \times \overline{\boldsymbol{H}}_n\right) \cdot \hat{\boldsymbol{z}}\, dx = \begin{cases} 1, & n = m \\ 0, & n \neq m \end{cases} \tag{S8}$$

is satisfied, where $\hat{\boldsymbol{z}}$ is the unit vector parallel to the axis of propagation of the guide. Neglecting radiation terms (which can be assumed to be negligible provided mode ovelap calculations below are performed at points sufficiently far, along the waveguide, from any waveguide inhomogeneities, for which radiating fields are negligible, and in view of (S8), the coefficients in (S7) are given by the mode overlap integrals

$$\begin{aligned} \boldsymbol{a}_m &= 0.5 \left(\int_{-\infty}^{\infty} \left(\boldsymbol{E} \times \overline{\boldsymbol{H}}_m\right) \cdot dx + \int_{-\infty}^{\infty} \left(\boldsymbol{H} \times \overline{\boldsymbol{E}}_m\right) \cdot dx \right) \\ \boldsymbol{b}_m &= 0.5 \left(\int_{-\infty}^{\infty} \left(\boldsymbol{E} \times \overline{\boldsymbol{H}}_m\right) \cdot dx - \int_{-\infty}^{\infty} \left(\boldsymbol{H} \times \overline{\boldsymbol{E}}_m\right) \cdot dx \right). \end{aligned} \tag{S9}$$

In most situations, it can be assumed that power is only travelling in the forward direction, implying that $b_m = 0$ and simplifying the expression for a_m:

$$\boldsymbol{a}_m = \int_{-\infty}^{\infty} \left(\boldsymbol{E} \times \overline{\boldsymbol{H}}_m\right) \cdot dx. \tag{S10}$$

Thus, the total power propagating in the forward $\hat{\boldsymbol{z}}$ direction along the guide is given by

$$P_m = 0.5 \left|\left|\boldsymbol{a}_m\right|\right|^2. \tag{S11}$$

SIV Optimization

The WGF-BIE solver described in this work can be used in conjunction with a suitable inverse design method to automatically design photonic devices—including grating couplers, in-plane power splitters with arbitrary split ratios, and wavelength demultiplexers—that meet target performance specifications. A fitness metric for a given device is specified by an appropriately selected objective function. At a high level, the objective function evaluates how well a candidate design fits the desired specifications on observables—such as efficiency for a certain frequency or frequencies, or within a certain bandwidth, etc.—and returns a single, real scalar as a result. For a properly defined objective function, a device which minimizes (or maximizes) this function is said to be the optimal device fitting the specified performance criteria. Thus, labeling by $\boldsymbol{p} = (p_1, \ldots, p_n)$ the parameters which fully specify the structure of a device design, the objective function f maps any (feasible) value of $\boldsymbol{p}$ to the real scalar $y = f(\boldsymbol{p})$.

For example, consider the power splitting device shown in Fig. 1a of the main text. The shape of the splitting device should be optimized in such a way that power incident on the input SIW (A) reaches the output SIW's (B and C) in the desired splitting ratio. In order to achieve this, a proper parametrization of the device geometry must be chosen which depends on the vector of optimization parameters $\boldsymbol{p}$ and an objective function must be selected—

which upon optimization yields the required functionality. The simple, but prototypical, objective function

$$f(\boldsymbol{p}) = (f_B(\boldsymbol{p}) - r_B)^2 + (f_C(\boldsymbol{p}) - r_C)^2$$

could be used in this case, where (r_B, r_C) represent the output power split ratio ($r_B + r_C = 1$)—so that, e.g. a 50/50 power splitter would would result for the parameter selection $(r_B, r_C) = (0.5, 0.5)$—and where f_B and f_C are functions which compute the ratio of power going out of outputs B and C respectively to incident power on input A.

This optimization setup gives rise to several advantages. Indeed, minimization of the objective function f by means of a suitable optimization algorithm should lead to values of the parameter vector $\boldsymbol{p}$ for which f_B approaches r_B and f_C approaches r_C. Furthermore, the closer that the final optimized device f_B and f_C approaches the ideal r_B and r_C values, the more energy makes it into one of the outputs B/C rather than getting scattered elsewhere—since $f_B + f_C$ would approach 1. Thus, this objective function also favors a device which is efficient in an absolute sense. Since the computation of f_B and f_c requires a full electromagnetic solution of the system at every iteration of the optimization, use of an efficient underlying electromagnetic solver can be greatly advantageous. In the context of the waveguide mode propagation considered in the previous two supplementary sections, finally, the functions f_B and f_C are given by the power flow in each one the output waveguides—which, on account of the exponential decay of the fields away from each one of the waveguides, can be computed by means of mode overlap integrals sufficiently far from the inhomogeneities and, on account of the exponential decay of the fields away from each one of the waveguides, the overlap integrals can be truncated in adequately small segments orthogonal to the waveguides instead of integrals between $-\infty$ and ∞.

In order to adequately explore the (highly complex) optimization landscape associated with the photonic-device optimization problems under consideration we utilize a sequence of steepest-descent optimizations, each one initiated at a randomly selected initial value (or "seed") for the parameter $\boldsymbol{p}$. (Strategies that could lead to improved selection of initial

seeds are the subject of on-going work.) Since typically many different initial seeds must be utilized to reach a local minimum sufficiently close to optimality (if not necessarily a global one), it is critical to utilize efficient algorithms for the evaluation of the electromagnetic solution as well as its gradient with respect to the (generally high-dimensional) parameter vector $\boldsymbol{p}$. The optimization times reported for the devices presented in this paper were the total time required for the optimization algorithm to converge for a given initial seed.

The steepest descent methods we use require evaluation of the "sensitivities" of a given function f—that is, the components of the gradient of the objective function

$$\nabla_{\boldsymbol{p}} f = \left(\frac{\partial f}{\partial p_1}, \ldots, \frac{\partial f}{\partial p_N} \right) \tag{S12}$$

with respect to $\boldsymbol{p} = (p_1, , \ldots, p_N)$ at any given value of $\boldsymbol{p}$. The simplest approach to gradient evaluation is based on finite-difference approximation

$$\begin{aligned} \nabla_{\boldsymbol{p}} f(\boldsymbol{p}) \approx \Bigg(& \frac{f(\boldsymbol{p} + h\boldsymbol{e}_1) - f(\boldsymbol{p})}{h}, \frac{f(\boldsymbol{p} + h\boldsymbol{e}_2) - f(\boldsymbol{p})}{h}, \\ & \ldots, \frac{f(\boldsymbol{p} + h\boldsymbol{e}_N) - f(\boldsymbol{p})}{h} \Bigg) \end{aligned} \tag{S13}$$

(with an adequately small value of h), where $\boldsymbol{e}_i$ denotes the unit vector in the i-th coordinate direction. Clearly, the approximate expression (S13) requires evaluation of $f(\boldsymbol{p})$ as well as $f(\boldsymbol{p} + h\boldsymbol{e}_i)$ for $i = 1, \ldots, N$, for a total of $N + 1$ objective function evaluations and, therefore, $N + 1$ full electromagnetic system solves for single-frequency problems. The cost grows to $k(N + 1)$ electromagnetic solves for optimization problems involving k different frequencies. The method (S13) for gradient evaluation can therefore be extremely expensive for parameter spaces of high dimensionality, to the point of becoming impractical for challenging optimization problems with potentially hundreds of coordinate directions in the $\boldsymbol{p}$-parameter space.

To address this problem in the next section we present a gradient evaluation method which, based on an application of the adjoint-method [S8, 9] in the context of integral-

equation-based Maxwell solvers, can produce all gradient components at a cost of merely $2k$ electromagnetic solves—regardless of the number N of optimization parameters—thus providing very significant gains for problems with large values of N.

SV Rapid gradient computation algorithm using the adjoint method

We present the proposed adjoint-based integral gradient method for an integral equation of the form

$$\mathcal{K}[\boldsymbol{\xi}] = \boldsymbol{b}, \quad \boldsymbol{\xi} \in X. \tag{S14}$$

Clearly, this equation can be interpreted in both the continuous and the discrete standpoints described in the boundary integral formulation section of the main text: by setting $X = X_r$, $\mathcal{K} = \mathcal{K}_r$ and $\boldsymbol{b} = \boldsymbol{b}_r$, we have $\boldsymbol{\xi} = \boldsymbol{\xi}_r$ for both cases, $r = c$ and $r = d$.

The objective function $f(\boldsymbol{p})$, which depends on the values of the fields u and its gradient at certain locations in a device, can be explicitly computed via evaluation of certain integrals involving Green's functions as well as the solution Φ^{scat} of the discretized linear system of integral equations presented in the main text. The derivation in this section utilizes the related function

$$g : \mathbb{R}^N \times X \to \mathbb{R}; \quad g = g(\boldsymbol{p}, \boldsymbol{\xi}). \tag{S15}$$

For given $\boldsymbol{p}$ and $\boldsymbol{\xi}$, the value $g(\boldsymbol{p}, \boldsymbol{\xi})$ of the functional g equals the value that would be obtained as the correct density Φ^{scat} (the solution of the discretized WGF-BIE problem) is substituted by the given density $\boldsymbol{\xi}$ everywhere in the expressions for the objective function f. Note that, in particular, we have $g(\boldsymbol{p}, \Phi^{\text{scat}}) = f(\boldsymbol{p})$.

First we will define a few additional preliminary notations followed by the derivation of the rapid gradient evaluation method.

A. Preliminaries

For reasons that should become clear in the present section, we consider X as a vector space over the real numbers, and we define the (real valued) inner product of (complex-valued) elements $\boldsymbol{a}, \boldsymbol{b} \in X$ by

$$< \boldsymbol{a}, \boldsymbol{b} >_{\mathrm{Re}} = \Re \left(\int_{\Gamma} \boldsymbol{a}(\boldsymbol{r}) \overline{\boldsymbol{b}(\boldsymbol{r})} d\boldsymbol{r} \right) \text{ or } < \boldsymbol{a}, \boldsymbol{b} >_{\mathrm{Re}} = \Re \left(\boldsymbol{a}^T \, \overline{\boldsymbol{b}} \right),$$

depending on whether $r = c$ or $r = d$, respectively. Here $\boldsymbol{x}^T$ and $\overline{\boldsymbol{x}}$ denote transposition and conjugation of the vector $\boldsymbol{x}$, respectively, and the right hand side in the second equation denotes the matrix product of the column vectors $\boldsymbol{a}^T$ and $\boldsymbol{b}$ viewed as single-row and single-column matrices $\boldsymbol{a}^T$ and $\overline{\boldsymbol{b}}$, respectively. (The somewhat unusual notational selections for the space X, which make it a vector space over the real numbers, although the vectors are themselves complex valued, and which uses a real scalar product, are useful below in this section. Note, for example, that, in the case $X = \mathbb{C}^N$, the resulting Hilbert space is equivalent to $\mathbb{R}^{2N}$ with its usual Euclidean scalar product.)

The adjoint $\mathcal{K}^\star$ of the operator $\mathcal{K}$, which is defined by the relation

$$< \mathcal{K}[\boldsymbol{a}], \boldsymbol{b} >_{\mathrm{Re}} = < \boldsymbol{a}, \mathcal{K}^\star[\boldsymbol{b}] >_{\mathrm{Re}} \tag{S16}$$

for all $\boldsymbol{a}$ and $\boldsymbol{b}$ (see e.g. [S10]), can be expressed, in the continuous case $r = c$, in terms of the complex conjugates of certain integral operators given by expressions similar to the operators presented in the formulation of the main text, but with $\boldsymbol{r}$ and $\boldsymbol{r'}$ interchanged in the argument of the Green's functions $G^{\pm}$. In the discrete case $r = d$, $\mathcal{K}$ is simply a matrix and $\mathcal{K}^\star$ is the complex conjugate of the transpose of $\mathcal{K}$. Note that two discrete versions of the adjoint gradient algorithm emerge at this point, which result as either

1. The adjoint gradient methodology is applied to the discrete problem ($r = d$); or

2. The adjoint gradient methodology is applied to the continuous problem ($r = c$), and

then the resulting *continuous* adjoint problem is discretized.

In this paper we have utilized the approach 1. Approach 2 conceptually facilitates the creation of accelerated [S11] formulations for the adjoint equation system since, indeed, an algorithm for the adjoint operator is obtained by interchanging the integration and observation variables in a previously existing algorithm for the direct operator.

In what follows we consider situations in which the interface set Γ depends on the design-parameter vector $\boldsymbol{p}$, $\Gamma = \Gamma_{\boldsymbol{p}}$, and consequently so do the operator $\mathcal{K}$ as well as the right-hand side $\boldsymbol{b}$ and the solution $\boldsymbol{\xi}$ of equation (S14); we will consequently write, for a given value of $\boldsymbol{p}$, $\mathcal{K} = \mathcal{K}_{\boldsymbol{p}}$, $\boldsymbol{b} = \boldsymbol{b}(\boldsymbol{p}) \in X$ and $\boldsymbol{\xi} = \xi(\boldsymbol{p}) \in X$. In order to determine the sensitivity of the objective with respect to variations of the design parameters, the gradient of the objective function f,

$$\nabla_p f(\boldsymbol{p}) = \nabla_p \Big(g(\boldsymbol{p}, \boldsymbol{\xi}(\boldsymbol{p})) \Big) = \left(\frac{\partial g(\boldsymbol{p}, \boldsymbol{\xi}(\boldsymbol{p}))}{\partial p_1}, \ldots, \frac{\partial g(\boldsymbol{p}, \boldsymbol{\xi}(\boldsymbol{p}))}{\partial p_N} \right),$$

must be computed. We seek to do this by utilizing the chain rule with respect to the variables $\boldsymbol{p}$ and $\boldsymbol{\xi}$.

An initial concern, in an attempt to use the chain rule in this context, could arise in connection with the complex-valued character of the $\boldsymbol{\xi}$ variables, as $g(\boldsymbol{p}, \boldsymbol{\xi})$ is real valued, and, therefore, not a complex-analytic function of $\boldsymbol{\xi}$. This difficulty could easily be solved, of course, simply by viewing $\boldsymbol{\xi}$ as a real valued vector of dimension $2N$ instead of an N-dimensional complex vector in the case $r = d$, or as a space of functions with values in $\mathbb{R}^2$ instead of a space of functions with values in $\mathbb{C}$, in the case r = c. But this notational burden can be avoided by utilizing in all cases the concept of Gateaux differentiation (a sort of directional derivative that uses a direction vector of arbitrary non-negative length): the Gateaux derivative of the function $g = g(\boldsymbol{p}, \boldsymbol{\xi})$ with respect to $\boldsymbol{\xi} \in X$ in the direction given

by a function $\boldsymbol{h} \in X$ is simply given by

$$\frac{\partial g}{\partial \boldsymbol{\xi}}(\boldsymbol{p}, \boldsymbol{\xi}; \boldsymbol{h}) = \lim_{\substack{\varepsilon \to 0 \\ \varepsilon \in \mathbb{R}}} \frac{g(\boldsymbol{p}, \boldsymbol{\xi} + \varepsilon \boldsymbol{h}) - g(\boldsymbol{p}, \boldsymbol{\xi})}{\varepsilon}. \tag{S17}$$

Not only does use of the Gateaux derivative allow us to bypass the aforementioned notational issues, but it also incorporates naturally the continuous case $r = c$.

Remark 1. Note that, as a function of $\boldsymbol{h}$, $\frac{\partial g}{\partial \boldsymbol{\xi}}(\boldsymbol{p}, \boldsymbol{\xi}; \boldsymbol{h})$ is a linear mapping from X to $\mathbb{R}^N$ over the set of real scalars. (If complex scalars are used, linearity only results when g is complex-analytic as a function of ξ—a condition that is never satisfied for the cost functionals g we need to utilize.)

Using the Gateaux-derivative notation (and under mild smoothness assumptions on the functions g and $\boldsymbol{\xi}$), the ℓ-th component of that gradient is given by

$$\frac{dg}{dp_\ell}(\boldsymbol{p}, \boldsymbol{\xi}) = \frac{\partial g}{\partial p_\ell}(\boldsymbol{p}, \boldsymbol{\xi}) + \frac{\partial g}{\partial \boldsymbol{\xi}}(\boldsymbol{p}, \boldsymbol{\xi}; \frac{\partial \boldsymbol{\xi}}{\partial p_\ell}). \tag{S18}$$

The necessary properties of smoothness of solutions $\boldsymbol{\xi}$ of integral equations with respect to boundary variations are studied, in closely related contexts, in [S12–14].

Remark 2. For sufficiently smooth functions g and for each fixed vector $\boldsymbol{p} \in \mathbb{R}^n$ and $\boldsymbol{\xi} \in X$ the Gateaux derivative (S17) can be expressed in terms of a scalar product:

$$\frac{\partial g}{\partial \boldsymbol{\xi}}(\boldsymbol{p}, \boldsymbol{\xi}; \boldsymbol{h}) = \langle G(\boldsymbol{p}, \boldsymbol{\xi}), \boldsymbol{h} \rangle_{\mathrm{Re}} \quad \text{for some vector} \quad G(\boldsymbol{p}, \boldsymbol{\xi}) \in X. \tag{S19}$$

In the finite-dimensional case ($r = d$) this follows from evaluation of the limit of the quotient of increments in (S17) by means of the chain rule with respect to the $2N$ dimensional vector given by the real and imaginary componens of the vector $\boldsymbol{\xi}$. In the infinite-dimensional case ($r = c$) a general version of this result follows from an application of the Lax-Milgram Lemma to the linear and continuous functional of $\boldsymbol{h}$ defined in (S17).)

B. Derivation

Recalling from the previous section that $\mathcal{K} = \mathcal{K}_{\boldsymbol{p}}$, and introducing the notation

$$R(\boldsymbol{p}, \boldsymbol{\xi}) = \mathcal{K}_{\boldsymbol{p}}[\boldsymbol{\xi}] - \boldsymbol{b}(\boldsymbol{p}), \tag{S20}$$

the linear relation (S14) may be re-expressed in the form

$$R(\boldsymbol{p}, \boldsymbol{\xi}) = 0.$$

Using (S18) we obtain

$$\frac{dR}{dp_\ell}(\boldsymbol{p}, \boldsymbol{\xi}) = \frac{\partial R}{\partial p_\ell}(\boldsymbol{p}, \boldsymbol{\xi}) + \frac{\partial R}{\partial \xi}(\boldsymbol{p}, \boldsymbol{\xi}; \frac{\partial \boldsymbol{\xi}}{\partial p_\ell}) = 0. \tag{S21}$$

But, in view of (S20) we have

$$\frac{\partial R}{\partial p_\ell}(\boldsymbol{p}) = \frac{\partial \mathcal{K}_{\boldsymbol{p}}}{\partial p_\ell}[\boldsymbol{\xi}] - \frac{\partial b}{\partial p_\ell}(\boldsymbol{p}) \tag{S22}$$

and

$$\frac{\partial R}{\partial \xi}(\boldsymbol{p}, \boldsymbol{\xi}; \frac{\partial \boldsymbol{\xi}}{\partial p_\ell}) = \lim_{\substack{\varepsilon \to 0 \\ \varepsilon \in \mathbb{R}}} \frac{\mathcal{K}_{\boldsymbol{p}}[\boldsymbol{\xi} + \varepsilon \frac{\partial \boldsymbol{\xi}}{\partial p_\ell}] - \mathcal{K}_{\boldsymbol{p}}[\boldsymbol{\xi}]}{\varepsilon} = \mathcal{K}_{\boldsymbol{p}} \left[\frac{\partial \boldsymbol{\xi}}{\partial p_\ell} \right], \tag{S23}$$

where the last identity results from the linear character of the operator $\mathcal{K}_{\boldsymbol{p}}$.

Suppressing the $\boldsymbol{p}$-dependence in the notation (with exception of the $\boldsymbol{p}$-dependence in the $\mathcal{K}_{\boldsymbol{p}}$), in view of (S22) and (S23) we may re-express (S21) in the form

$$\frac{dR}{dp_\ell} = \frac{\partial \mathcal{K}_{\boldsymbol{p}}}{\partial p_\ell}[\boldsymbol{\xi}] - \frac{\partial \boldsymbol{b}}{\partial p_\ell} + \mathcal{K}_{\boldsymbol{p}} \left[\frac{\partial \boldsymbol{\xi}}{\partial p_\ell} \right] = 0. \tag{S24}$$

Taking the inner product of $\frac{dR}{dp_\ell}$ with an arbitrary (generally complex-valued) $\boldsymbol{\lambda} \in X$ yields

$$\begin{aligned}\left\langle \boldsymbol{\lambda}, \frac{dR}{dp_\ell} \right\rangle_{\mathrm{Re}} = &\left\langle \boldsymbol{\lambda}, \frac{\partial \mathcal{K}_{\boldsymbol{p}}}{\partial p_\ell}[\boldsymbol{\xi}](\boldsymbol{r}) - \frac{\partial \boldsymbol{b}}{\partial p_\ell} \right\rangle_{\mathrm{Re}} \\ &+ \left\langle \boldsymbol{\lambda}, \mathcal{K}_{\boldsymbol{p}} \left[\frac{\partial \boldsymbol{\xi}}{\partial p_\ell} \right] \right\rangle_{\mathrm{Re}} = 0.\end{aligned} \tag{S25}$$

This expression may be viewed as a linear combination, with real coefficients equal to the real and imaginary parts of $\boldsymbol{\lambda}(\boldsymbol{r})$, of all the equations obtained, one for each value of $\boldsymbol{r}$, from the real and imaginary parts of equation (S24). The power of the adjoint method lies on its subtraction of the resulting "linear-combination equation" from the expression (S18) for the gradient components, with a selection of coefficients $\boldsymbol{\lambda}$ such that the unknown term $\frac{\partial \boldsymbol{\xi}}{\partial p_\ell}$ does not appear the resulting expression. As discussed in what follows, the necessary coefficients can be obtained as the solution of a certain adjoint equation, and the gradient of the objective function can thus be produced on the basis of just two system solves—at significant savings when compared with the direct gradient evaluation inherent in (S13), which requires a total of $N + 1$ system solves.

To pursue this idea we note that, in view of (S16), we have $< \boldsymbol{\lambda}, \mathcal{K}_{\boldsymbol{p}}[\frac{\partial \boldsymbol{\xi}}{\partial \boldsymbol{p}}] >_{\mathrm{Re}} = < \mathcal{K}^\star_{\boldsymbol{p}}[\boldsymbol{\lambda}], \frac{\partial \boldsymbol{\xi}}{\partial \boldsymbol{p}} >_{\mathrm{Re}}$, (in other words, the linear-combination equation may be expressed as a linear combination of the real and imaginary parts of $\frac{\partial \boldsymbol{\xi}}{\partial \boldsymbol{p}}$) and, therefore,

$$\left\langle \boldsymbol{\lambda}, \frac{dR}{dp_\ell} \right\rangle_{\mathrm{Re}} = \left\langle \boldsymbol{\lambda}, \frac{\partial \mathcal{K}_{\boldsymbol{p}}}{\partial p_\ell}[\boldsymbol{\xi}] - \frac{\partial \boldsymbol{b}}{\partial p_\ell} \right\rangle_{\mathrm{Re}} + \left\langle \mathcal{K}^\star_{\boldsymbol{p}}[\boldsymbol{\lambda}], \frac{\partial \boldsymbol{\xi}}{\partial p_\ell} \right\rangle_{\mathrm{Re}} = 0,$$

which, when subtracted from (S18) yields

$$\begin{aligned}\frac{dg(\boldsymbol{p}, \boldsymbol{\xi})}{dp_\ell} = &\left\{ \frac{\partial g}{\partial p_\ell} - \left\langle \boldsymbol{\lambda}, \frac{\partial \mathcal{K}_{\boldsymbol{p}}}{\partial p_\ell}[\boldsymbol{\xi}] - \frac{\partial \boldsymbol{b}}{\partial p_\ell} \right\rangle_{\mathrm{Re}} \right\} \\ &+ \left\{ \frac{\partial g}{\partial \xi}(\boldsymbol{p}, \boldsymbol{\xi}; \frac{\partial \boldsymbol{\xi}}{\partial p_\ell}) - \left\langle \mathcal{K}^\star_{\boldsymbol{p}}[\boldsymbol{\lambda}], \frac{\partial \boldsymbol{\xi}}{\partial p_\ell} \right\rangle_{\mathrm{Re}} \right\}.\end{aligned} \tag{S26}$$

We now seek $\boldsymbol{\lambda} \in X$ such that the second term in curly brackets on the RHS of this equation vanishes for all possible functions $\frac{\partial \boldsymbol{\xi}}{\partial p_\ell}$—since, upon substitution of such an element $\boldsymbol{\lambda}$ in

equation (S26) we obtain an expression for $\frac{dg(\boldsymbol{p},\boldsymbol{\xi})}{dp_\ell}$ which does not depend on $\frac{\partial \boldsymbol{\xi}}{\partial p_\ell}$, as desired. Using (S19) we see that the needed $\boldsymbol{\lambda} \in X$ can be obtained as the solution of the "adjoint equation"

$$\mathcal{K}^\star_{\boldsymbol{p}}[\boldsymbol{\lambda}] = G(\boldsymbol{p}, \boldsymbol{\xi}); \tag{S27}$$

using such $\lambda \in X$, the desired expression for the gradient $\frac{dg(\boldsymbol{p},\boldsymbol{\xi})}{dp_\ell}$ is obtained:

$$\frac{dg(\boldsymbol{p},\boldsymbol{\xi})}{dp_\ell} = \left\{ \frac{\partial g}{\partial p_\ell} - \left\langle \boldsymbol{\lambda}, \frac{\partial \mathcal{K}_{\boldsymbol{p}}}{\partial p_\ell}[\boldsymbol{\xi}] - \frac{\partial \boldsymbol{b}}{\partial p_\ell} \right\rangle_{\mathrm{Re}} \right\}, \tag{S28}$$

in which, since $\boldsymbol{\xi}$ is kept fixed as the derivative $\frac{\partial \mathcal{K}_{\boldsymbol{p}}}{\partial p_\ell}[\boldsymbol{\xi}]$ on the right-hand side is evaluated, the dependence on $\frac{\partial \boldsymbol{\xi}}{\partial p_\ell}$ has been eliminated.

SVI Numerical Implementation

This section provides details on the numerical implementation of the WGF-BIE solver for modeling nanophotonic devices.

Boundary Representation by Parametrized Segments

The boundary integral formulation described in the main text requires integration of various types of singular integrals over the boundary set Γ. In the proposed integral-equation solver, the curve Γ is expressed as a union of parametrized curved segments, each one of which is represented by a vector function of the form

$$\boldsymbol{r} = \boldsymbol{\gamma}(t), \quad 0 \le t \le 2\pi. \tag{S29}$$

(The parametrization interval $[0, 2\pi]$ we use is particularly convenient in the context of the spectral methods described in the following section.) The spectral character of the algorithms leads to high accuracies in short computing times, but any available integral-

equation method, such as e.g. the Method of Moments [S15], could be used in this context instead.

Fig. 1a in the main text demonstrates, in the context of a power splitter device, the representation of the boundary set Γ as a union of parametrized segments. In this case, the parametrization was obtained from an overall B-Spline representation [S16].

Two aspects of the boundary discretization are important to note:

1. Any arbitrary parametrization may be used to represent each boundary. In this work, we utilize B-spline curves to represent the boundaries of the power splitter and waveguide taper devices and standard line segments to represent the boundaries of the grating coupler device.

2. We use a non-uniform Nyström mesh (described in the following subsection) which, relying on use of certain changes of variables, produces discretizations that contain higher density of unknowns in a neighborhood of corners and thus give rise to high-order accuracy throughout the whole simulation domain in the context of the multi-segment geometry-representation paradigm used, including at the corners themselves (where the fields exhibit singular behavior [S17]).

Numerical Quadrature and Discretization

The proposed numerical "Nyström" solver produces a highly-accurate discrete version of the windowed integral equations. This discrete set of equations is constructed on the basis of the parametrized-segment representation described in the previous section: discretizing each segment in the geometry (according to a strategy described in what follows) two discretization vectors are formed, namely, a "point-vector" $\boldsymbol{r}'_d \in \mathbb{R}^{2N}$ (a two-dimensonal vector for each one of the N discretization points) containing all of the discretization points contributed by all parametrized segments, and a corresponding "unknowns-vector" $\boldsymbol{\xi}_d \in \mathbb{C}^{2N}$ containing the (unknown) numerical values of the integral-equation solution, one at each point represented

in the point-vector $\boldsymbol{r}'_d$. Given this information, together with some representation for the normal and the element of length along the curve, the values of the integrals on the left-hand side of the windowed boundary integral formulation can be obtained by numerical quadrature for any point $\boldsymbol{r}$ on the boundary curves. The numerical evaluation of such integrals at each one of the spatial points represented by a certain collocation point-vector $\boldsymbol{r}_d$ (which is in fact taken to equal $\boldsymbol{r}'_d$, so that the set of evaluation points coincides with the set of integration points), can be expresseed as a matrix-vector product. Upon additional numerical evaluations for the right-hand integrals at the point-vector $\boldsymbol{r}_d$, along the lines described in [S4], a discrete linear system of equations of the form $\mathcal{K}_d\boldsymbol{\xi}_d = b_d$ is obtained. The methods utilized for the treatment of each one of these integration problems are described in what follows.

Each one of the necessary integrals over Γ is expressed as a sum of integrals over each one of the parametrized segments introduced in the previous section. To obtain equations for the unknown densities, it thus suffices to evaluate (and suitably combine) each one of these segment-integrals for each $\boldsymbol{r}$ in an appropriately selected set of "collocation points" contained in Γ—each one of which gives rise to an equation for the unknowns in the problem. The unknowns themselves are approximate numerical values of the continuous integral-equation unknowns $\Phi^{\text{scat}} = (\phi^{\text{scat}}, \psi^{\text{scat}})^T$ at points $\boldsymbol{r}'$ on a suitably selected set of "quadrature points", which as mentioned above, are taken to coincide with the collocation points—and which can be used to produce all the necessary integrals at each one of the selected collocation points. (In cases in which two segment boundaries are very close to each other, a refinement procedure is used in which the set of integration points is refined as needed, but without increasing the density of unknowns, to adequately resolve the near-singularity of the Green function. The additional values of the unknown density needed for integration are obtained by interpolation on the basis of the underlying Chebyshev expansion.) The numerical method can then be set up easily by either forming the corresponding matrix $\mathcal{K}_d$ that embodies the integration and collocation process mentioned above, or by alternatively producing an

algorithm (the "forward-map" procedure) that evaluates the product of $\mathcal{K}_d$ and any given vector in the solution space, but without necessarily forming the matrix A. In the first approach, which we used in this paper, the problem can be ultimately solved by means of Gaussian elimination. The second approach can lead to significantly reduced computational times if an appropriate acceleration techique is utilized (see [S11] and references therein), but this line of development lies beyond the scope and needs of the present paper.

The prescriptions for the quadrature method are completed by providing a quadrature method for integration in each one of the integration segments. Using the available parametrization for a given segment, an integral over the segment is at first transformed into an integral over the interval $[0, 2\pi]$. In order to reduce this integration problem to numerical quadrature with high-order accuracy, our quadrature algorithm incorporates, in a novel fashion, the polynomial changes of variables that are used in [S18] for treatment of corner-singularities in the integral-equation solutions. As in that reference, the changes of variables are used here to treat singularities at corners. But, additionally, even for smooth regions, the changes of variables are used by our algorithms to induce periodicity in the integrand, and, thus, to give rise to high-order (spectral) accuracy as the trapezoidal-rule and spectral integration for the logarithmic kernel are used as indicated in [S17, Sec. 3.5].

Localized Matrix Updates

At the lowest level, the system matrix for the WGF-BIE formulation maps each unknown as a source point to all of the other unknowns as target points. Normally, if we perturb the structure that we wish to simulate, we would have to rebuild the complete system matrix. What if we only perturb the locations of a few nodes, however, leaving the majority of the nodes untouched? In this case, only the entries which correspond to the interactions of either a changed node as a source to all of the other nodes or any of the unchanged nodes as a source to a changed node as a target would change and all of the other interactions between unchanged nodes *would remain completely unchanged*. This implies that if we use

a *local* parametrization for each curve representing our device, such that perturbing one optimization parameter affects nodes corresponding only to one (or a few) of the curves, then we can very efficiently compute the new system matrix for each perturbation by simply *updating* the original, unperturbed system matrix. This allows very rapid computation of the $\frac{\partial \mathcal{K}_d}{\partial p_i}$ matrices required for implementing the adjoint method via matrix updates and finite differences.

References

(S1) Neumann, E.-G. *Single-mode fibers: fundamentals*; Springer, 2013; Vol. 57.

(S2) Svelto, O.; Hanna, D. C. *Principles of lasers*; Springer, 1998; Vol. 4.

(S3) Magnanini, R.; Santosa, F. Scattering in a 2-D Optical Waveguide. *Chapman and Hall Crc Research Notes in Mathematics* **2000**,

(S4) Bruno, O. P.; Garza, E.; Pérez-Arancibia, C. Windowed Green function method for nonuniform open-waveguide problems. *IEEE Transactions on Antennas and Propagation* **2017**, *65*, 4684–4692.

(S5) Nosich, A. I. Radiation conditions, limiting absorption principle, and general relations in open waveguide scattering. *Journal of electromagnetic waves and applications* **1994**, *8*, 329–353.

(S6) Snyder, A. W.; Love, J. *Optical waveguide theory*; Springer Science & Business Media, 2012.

(S7) Ref. 6, eqs. (11-2), (11-9).

(S8) Giles, M. B.; Pierce, N. A. An introduction to the adjoint approach to design. *Flow, turbulence and combustion* **2000**, *65*, 393–415.

(S9) Jameson, A. Optimum aerodynamic design using CFD and control theory. 12th Computational Fluid Dynamics Conference. 1995; p 1729.

(S10) Dunford, N.; Schwartz, J. T. *Linear operators part I: general theory*; Interscience publishers New York, 1958; Vol. 7.

(S11) Bruno, O. P.; Kunyansky, L. A. A fast, high-order algorithm for the solution of surface scattering problems: basic implementation, tests, and applications. *Journal of Computational Physics* **2001**, *169*, 80–110.

(S12) Bruno, O. P.; Reitich, F. Numerical solution of diffraction problems: a method of variation of boundaries. *JOSA A* **1993**, *10*, 1168–1175.

(S13) Nicholls, D. P.; Reitich, F. Stability of high-order perturbative methods for the computation of Dirichlet–Neumann operators. *Journal of Computational Physics* **2001**, *170*, 276–298.

(S14) Bruno, O. P.; Laurence, P. Existence of three-dimensional toroidal MHD equilibria with nonconstant pressure. *Communications on pure and applied mathematics* **1996**, *49*, 717–764.

(S15) Harrington, R. F. *Field computation by moment methods*; Wiley-IEEE Press, 1993.

(S16) De Boor, C. *A practical guide to splines*; springer-verlag New York, 1978; Vol. 27.

(S17) Colton, D.; Kress, R. *Inverse Acoustic and Electromagnetic Scattering Theory*, 3rd ed.; Springer: New York, 2013.

(S18) Ref. 17, eq. (3.104).