

RESEARCH ARTICLE OPEN ACCESS

Efficient First-Principles Inverse Design of Nanolasers

Beñat Martinez de Aguirre Jokisch^{1,2} | Alexander Cerjan³ | Rasmus Ellebæk Christiansen^{1,2} | Jesper Mørk^{2,4} | Ole Sigmund^{1,2} | Steven G. Johnson⁵

¹Department of Civil and Mechanical Engineering, Technical University of Denmark, Kongens Lyngby, Denmark | ²NanoPhoton – Center for Nanophotonics, Technical University of Denmark, Kongens Lyngby, Denmark | ³Center for Integrated Nanotechnologies, Sandia National Laboratories, Albuquerque, NM, USA | ⁴Department of Electrical and Photonics Engineering, Technical University of Denmark, Kongens Lyngby, Denmark | ⁵Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, USA

Correspondence: Beñat Martinez de Aguirre Jokisch (bmdaj@dtu.dk) | Rasmus Ellebæk Christiansen (raelch@dtu.dk)

Received: 25 June 2025 | **Revised:** 8 March 2026 | **Accepted:** 9 March 2026

Keywords: cavity | inverse design | laser | topology optimization

ABSTRACT

We develop and demonstrate a first-principles approach, based on the nonlinear Maxwell–Bloch equations and steady-state *ab-initio* laser theory (SALT), for inverse design of nanostructured lasers, incorporating spatial hole-burning corrections, threshold effects, out-coupling efficiency, and gain diffusion. The resulting figure of merit exploits the high-Q regime of optimized laser cavities to perturbatively simplify the nonlinear model to a single *linear* “reciprocal” Maxwell solve. The consequences for laser-cavity design, and in particular the strong dependence on the nature of the gain region, are demonstrated using topology optimization of both 2d and full 3d geometries.

1 | Introduction

In this paper, we present a new formulation for efficient inverse design of nanolasers (Figure 1) to maximize efficiency ($\sim$ output power/pump power), taking nonlinear and coupling effects into account, and demonstrate 2d and 3d semiconductor-laser topology optimization (TopOpt, Section 2.5). We exploit the fact that optimization evolves the lasing cavity to a high-Q ($\gtrsim 100$, long-lifetime) regime to perturbatively extract hole-burning effects, thresholds, output coupling, and other relevant phenomena from a *single linear* “reciprocal” scattering solve excited from the output port for a passive cavity. Our approach builds on the “SALT” (steady-state *ab initio* laser theory) formulation of lasing [1, 2] (Section 2.1), which exactly solves the nonlinear Maxwell–Bloch equations for a single-mode lasing steady state, including all nonlinear gain and saturation effects in the Maxwell–Bloch model, but we are able to eliminate expensive nonlinear and linear eigensolves using a new synthesis of SALT, perturbation theory (extending single-pole approximation (SPA)-SALT theory [2], Section 2.2), and temporal coupled-mode theory

(TCMT [3], Section 2.3). Our resulting figure of merit (FOM) is a simple ratio $\sim (\int_{\text{gain}} |\mathbf{E}|^2 d\Omega)^3 / \int_{\text{gain}} |\mathbf{E}|^4 d\Omega$ (Equation (14) in Section 2.3.3) computed from only the linear-model electric field $\mathbf{E}$. We show that this FOM reduces to a more conventional cavity LDOS-like optimization—power emitted into an output channel by a dipole—in the limit of a *point-like* gain region (Section 2.3.4), but leads to very different (more delocalized) cavity designs and performance for a *distributed* gain region. We demonstrate TopOpt of this FOM with both 2d and 3d semiconductor-on-insulator designs, including manufacturing constraints, and investigate the effect of gain-region diameter (Section 3.1). We also compare to a heuristic FOM $\sim \int_{\text{gain}} |\mathbf{E}|^2 d\Omega$ (Section 2.3.4), and show that this leads to suboptimal performance (Section 3.1). Furthermore, we show how to incorporate gain-diffusion effects [4] into TopOpt (e.g., for semiconductor lasers in a free-carrier approximation) at the cost of only two additional damped-diffusion solves (Section 2.4), and show that diffusion favors a disconnected topology of the gain region (Section 3.2). We believe that our results not only demonstrate the importance (and surprising ease) of incorporating sophisticated

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Laser & Photonics Reviews* published by Wiley-VCH GmbH

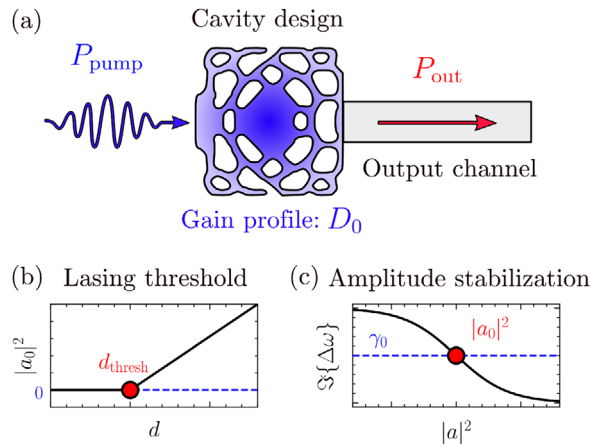


FIGURE 1 | (a) Schematic nanolaser: pump power P_{pump} excites a gain profile D_0 and a lasing mode that outputs power P_{out} into a channel (waveguide), for a cavity (from Section 3.1) that maximizes efficiency $\sim P_{\text{out}}/P_{\text{pump}}$. (b) Lasing amplitude $|a_0|^2$ ($\sim$ energy) vs. pump strength d , lasing above a threshold d_{thresh} . (c) Change $\Im\{\Delta\omega\}$ in modal gain/loss depends nonlinearly on amplitude $|a|^2$, and stabilizes at $|a_0|^2$, where $\Im\{\Delta\omega\} = \text{passive cavity loss } \gamma_0$.

nonlinear lasing models into the inverse-design process, but also provide a foundation for future extensions, e.g., to include an explicit model of the pumping process, as discussed in Section 4.

The inverse design of nanolasers has conventionally been based on improving generic FOMs that are *related* to lasing, without directly modeling the laser physics or accounting for the spatial distribution of the emitters [5], such as maximizing the cavity quality factor Q ($\sim$ lifetime/period [3]). Another such FOM is the local density of states (LDOS), which is proportional to Q divided by a measure of modal volume (the “Purcell factor”) and is more precisely the total power expended by a dipole source [6]; this is related to enhancement of light–matter interactions (e.g., spontaneous emission rate) at a single point [7–10]. For example, high Q factors lead to lower lasing thresholds [11] (Section 2.2). Nonetheless, and as our results also demonstrate, in nanolaser design a high Q alone is insufficient: while strong temporal confinement (high Q) is crucial, the photons emitted by the gain medium must also be efficiently extracted and coupled into an output channel. This concept has been explored in the context of inverse design of light emission, where the optical extraction efficiency [12–20] —the (linear-model) output power of some source into some channel(s)—is considered as the optimization FOM. However, for laser design, simple metrics such as the Q , LDOS, or extraction efficiency do not account for the precise distribution of the gain medium and its overlap with the optical field [5] or nonlinear lasing effects (e.g., gain, saturation). Below, we show that an LDOS-like FOM only becomes equivalent to a more complete lasing model in the limiting case of a point-like gain region (as shown in Section 2.3.4). The LDOS is also relevant to other physical phenomena such as sensing, and there have been several research efforts targeting this FOM using inverse-design frameworks like TopOpt [21–25], including experimental realization [24], along with other methods targeting high Q [26]. (TopOpt, reviewed in Section 2.5, is a gradient-based inverse design framework that finds freeform designs via large-scale optimization where “every pixel” is a degree of

freedom, and has been successfully applied to many photonic problems [27, 28].) However, it is known that the result of cavity optimization can be greatly altered if one targets an FOM that is tailored to a phenomenon very different from spontaneous emission at a point, such as Raman scattering [29–31], volume-averaged emission [16], scintillation effects [32], or harmonic generation [33, 34].

A key enabling factor for first-principles nanolaser design is the recent development of SALT, reviewed in Section 2.1 [1, 2], which formulates a tractable model of a full nonlinear lasing problem by transforming the time-domain nonlinear Maxwell–Bloch equations [35, 36] into a set of self-consistent, time-independent equations for the lasing steady state. SALT has provided a foundation for many subsequent developments, such as improvements in the theory of laser amplifiers [37], complicated gain media with carrier diffusion [4], quantum limits on the laser linewidth [38], and random lasers [39]. However, the fact that SALT is based on a nonlinear eigenproblem (arising when a resonance lases) makes it more challenging for inverse design than linear scattering: not only does it have all of the pole-tracking and potential non-differentiability of linear eigenvalue/resonance optimization [22], but when a pole crosses the real- ω axis one must also “turn on” a nonlinear solver [40]. Merely identifying the lasing pole in a differentiable way would be difficult without a good starting guess (perhaps via LDOS optimization), and the nonlinear effects incur additional computational complexity [40]. Although these issues may be tractable in principle (e.g., differentiating the nonlinear solve via implicit adjoint methods [41]; such methods have been used to optimize nonlinear steady-state transmission [42]), it would be better to find a more computationally efficient formulation. The reason such an improvement should be possible is the fact that an *optimized* nanolaser is quite special: to be high-efficiency and/or low threshold, the cavity will have a high Q , and it is known that SALT can be simplified by perturbative analysis in this regime [2] (empirically found to be accurate for passive cavity resonances with $Q \gtrsim 100$ [43] and numerically validated in Appendix D), although the previous perturbative “SPA-SALT” techniques were not devised with optimization in mind. Here, we show that specifically targeting efficient inverse design allows us to obtain a computationally efficient and simple FOM without sacrificing first-principles physics.

2 | Theoretical Framework

In this section, we introduce the theoretical framework to model nanolasers using SALT in the perturbative single-mode high- Q regime, and combining it with TCMT, we derive a FOM for nanolaser optimization that can be evaluated with a single reciprocal scattering solve. First, we review laser modeling via SALT (Section 2.1), which requires one to solve a system of frequency-domain nonlinear equations to calculate laser properties. Then, we review SPA-SALT (Section 2.2) an approximation for the single-mode high- Q cavity limit, where perturbation theory is used to derive analytical expressions for laser properties in terms of the passive cavity (no gain) mode solution. In this limit, the equations simplify to a single eigenvalue problem. Next, we use TCMT (Section 2.3) to derive a FOM (Equation (14)) that accounts for the out-coupling into a desired channel, and can be computed with a single reciprocal solve, where instead

of considering the emission from the gain medium one excites an output channel and evaluates the linear-Maxwell electromagnetic fields in the cavity. We then compare the FOM to a heuristic generalization of the LDOS for a distributed gain medium, and show that both FOMs are equivalent in the single-emitter limit (Section 2.3.4). Finally, we define the topology optimization framework (Section 2.5) to set up an optimization problem that can be solved to maximize the FOM.

2.1 | Review of Steady-State Ab-initio Laser Theory (SALT)

In a laser, an external pump excites a gain medium, which then interacts with the optical cavity mode, and emits photons into an output mode through stimulated emission. This is illustrated in Figure 1a for an example of an inverse designed nanolaser. The gain medium located in the cavity is modeled as an ensemble of two-level atoms interacting with the optical cavity field, which can be described by means of the Maxwell-Bloch equations [1, 35, 36]

$$\begin{aligned} \nabla \times \nabla \times \mathbf{E}^+ - \frac{\epsilon_c}{c^2} \ddot{\mathbf{E}}^+ &= \frac{1}{c^2 \epsilon_0} \ddot{\mathbf{P}}^+, \\ \dot{\mathbf{P}}^+ &= (-i\omega_a - \gamma_\perp) \mathbf{P}^+ + \frac{g}{i\hbar} \mathbf{E}^+ D, \\ \dot{D} &= \gamma_\parallel (D_0 - D) - \frac{2}{i\hbar} [\mathbf{E}^+ \cdot (\mathbf{P}^+)^* - \mathbf{P}^+ \cdot (\mathbf{E}^+)^*], \end{aligned} \quad (1)$$

where i is the imaginary unit, $\hbar$ is the reduced Planck constant, g is a dipole matrix element, $\mathbf{E}^+(\mathbf{r}, t)$ and $\mathbf{P}^+(\mathbf{r}, t)$ are the positive frequency components of the electric field and polarizations respectively, $\epsilon_c(\mathbf{r})$ is the complex relative permittivity of the cavity, ϵ_0 is the vacuum permittivity, $D(\mathbf{r}, t)$ is the population inversion, $D_0(\mathbf{r})$ is the gain profile, which is established through external pumping, ω_a is the atomic transition frequency, $\gamma_\perp$ is the gain width (polarization dephasing rate), and $\gamma_\parallel$ is the population relaxation rate. Note that these equations rely on the rotating-wave approximation (RWA) [2].

For semiconductor lasers, the population relaxation rate will be much lower than the atomic transition frequency and the dephasing rate ($\omega_a, \gamma_\perp \gg \gamma_\parallel$), which leads to the stationary-inversion approximation (SIA) $\dot{D} \sim 0$ [2]. One can then derive the coupled nonlinear SALT differential equations [2]

$$\left[(\nabla \times \nabla \times) - \left(\epsilon_c(\mathbf{r}) + \frac{\gamma_\perp D(\mathbf{r})}{\tilde{\omega}_i - \omega_a + i\gamma_\perp} \right) \left(\frac{\tilde{\omega}_i}{c} \right)^2 \right] \mathbf{E}_i(\mathbf{r}) = 0, \quad (2)$$

where ω_i and $\mathbf{E}_i$ are the complex-valued frequency and mode profile of the i -th cavity mode, and the expression for the population inversion is given by

$$D(\mathbf{r}) = \left[1 + \sum_{j=1}^N \Gamma_j e_c^{-2} |\mathbf{E}_j(\mathbf{r})|^2 \right]^{-1} D_0(\mathbf{r}) d, \quad (3)$$

where d is a scalar-valued pumping strength that we have separated from the spatial pumping profile $D_0(\mathbf{r})$, $\Gamma_j = \gamma_\perp^2 / [\gamma_\perp^2 + (\omega_j - \omega_a)^2]$ is a Lorentzian gain curve evaluated at the frequency ω_j for a gain medium emitting at the wavenumber ω_a , and e_c is a non-dimensionalization scaling for the fields [2].

Note that this framework does not explicitly model the physical pumping process: $D_0 d$ is treated here as a fixed input. As discussed in Section 4, an opportunity for future work is to obtain D_0 itself by solving an auxiliary system of equations, e.g., another Maxwell solve for optical pumping. Also not explicitly included here is the pump power to reach transparency ($d = 0$). Although this could be larger than the additional $d > 0$ pump power required to reach/exceed threshold for high Q and low lasing powers, the power to reach transparency is mostly a property of the materials and the pumping process, independent of the lasing mode, so it is not our primary concern in this paper.

2.2 | Single-Pole Threshold Approximation

The expression in Equation (2) simplifies when considering a single high- Q lasing mode L ($i = j = L$) that is aligned to the atomic transition $\omega_a = \omega_L$ of the emitters in the gain medium. This approximation, which is accurate for a sufficiently high Q (e.g., $Q \gtrsim 100$ in previous work [43] and numerical validation in Appendix D) is also known as the single-pole approximation (SPA-) SALT equation [2]

$$\left[(\nabla \times \nabla \times) - (\epsilon_c(\mathbf{r}) - i\Delta\epsilon_\mathfrak{S}(\mathbf{r})) \left(\frac{\omega_L}{c} \right)^2 \right] \mathbf{E}_L(\mathbf{r}) = 0. \quad (4)$$

where the change in the dielectric permittivity is given by

$$\Delta\epsilon_\mathfrak{S}(\mathbf{r}) = \frac{D_0(\mathbf{r})d}{1 + e_c^{-2} |\mathbf{E}_L(\mathbf{r})|^2}. \quad (5)$$

Equation (4) still involves solving one nonlinear eigenvalue problem. Nevertheless, the expression can be further simplified by perturbatively reducing the equation to linear form. For a high- Q cavity, operating at low powers near threshold, the gain required to balance cavity loss is small. More precisely, in this regime the change in the imaginary part of the dielectric permittivity induced by the active medium will be small ($\propto 1/Q$) with respect to the passive cavity permittivity ($\Delta\epsilon_\mathfrak{S} \ll \epsilon_c$), for a high- Q passive cavity mode with a complex angular frequency $\tilde{\omega}_1 = \omega_1 - i\gamma_1$, where the quality factor is defined as $Q = \omega_1/(2\gamma_1)$. We can then describe the shift in the cavity-mode frequency by using first-order perturbation theory [44]

$$\Delta\tilde{\omega}_1 = i \frac{\omega_1}{2} \frac{\int_\Omega \Delta\epsilon_\mathfrak{S}(\mathbf{r}) |\mathbf{e}_1(\mathbf{r})|^2 d\Omega}{\int_\Omega \epsilon_c(\mathbf{r}) |\mathbf{e}_1(\mathbf{r})|^2 d\Omega}, \quad (6)$$

where $\mathbf{e}_1$ is the mode profile of the passive cavity. In the high- Q cavity limit, the lasing mode $\mathbf{E}_L$ is approximately proportional to the cavity mode [2] within a $\mathcal{O}(1/\sqrt{Q})$ error (proportional to the square root of the radiated power [3]). The change in permittivity can then be expressed as

$$\Delta\epsilon_\mathfrak{S}(\mathbf{r}) \approx \frac{D_0(\mathbf{r})d}{1 + |a_1|^2 (e_c^{-2} |\mathbf{e}_1(\mathbf{r})|^2 / \int_\Omega \epsilon_c(\mathbf{r}) |\mathbf{e}_1(\mathbf{r})|^2 d\Omega)}, \quad (7)$$

where a_1 is the amplitude of the lasing mode and the lasing intensity $|a_1|^2$ has units of energy. Note that since $\Delta\epsilon_\mathfrak{S} \in \mathbb{R}$ this yields a purely imaginary shift in the cavity-mode frequency, which can be used to counteract the losses of the passive cavity, so that the cavity mode starts lasing.

In the low-power regime, for a small amplitude $|a_1|^2$ of the lasing mode, one can expand Equation (7) in a Taylor series around $|a_1|^2 = 0$. As shown in Figure 1b the amplitude of the lasing mode is zero at threshold, and the system will lase for a pumping strength above threshold $d \geq d_{\text{thresh}}$. Notice, that we are here considering the approximation where we neglect the spontaneous emission into the lasing mode, which will lead to non-zero power in the cavity mode below the lasing threshold [11]. In the zero-amplitude limit, the change in permittivity is linearly related to the pump strength: $\Delta\epsilon_3(\mathbf{r}) = d D_0(\mathbf{r})$. As shown in Figure 1c, for lasing to occur, the imaginary frequency shift caused by the gain, as expressed in Equation (6), must balance the losses of the passive cavity

$$\Im\{\Delta\tilde{\omega}_1\} - \gamma_1 = 0. \quad (8)$$

This condition is met at threshold ($d = d_{\text{thresh}}$), giving an expression for the pumping strength needed to reach threshold

$$d_{\text{thresh}} = \frac{1}{Q} \frac{\int_{\Omega} \epsilon_c(\mathbf{r}) |\mathbf{e}_1(\mathbf{r})|^2 d\Omega}{\int_{\Omega} D_0(\mathbf{r}) |\mathbf{e}_1(\mathbf{r})|^2 d\Omega}. \quad (9)$$

As would be expected, the pumping strength needed to reach the lasing threshold is inversely proportional to Q and to the energy confinement in the active medium, which leads to stronger light-matter interaction. In Appendix A, we show how the same result can be recovered using SPA-SALT. Note that in the single-emitter limit where $D_0(\mathbf{r}) \propto \delta(\mathbf{r} - \mathbf{r}')$, Equation (9) simplifies to $d_{\text{thresh}} = V/Q$, where V is the typical “mode volume” of the resonance, recovering an expression inversely proportional to the Purcell enhancement and the LDOS [7–9]; we comment further on this regime in Section 2.3.4.

Once the pumping strength crosses the lasing threshold, the cavity-mode will start to lase with an amplitude a_1 , as depicted in Figure 1b. Considering both the zeroth-order and the first-order correction in the small amplitude expansion of Equation (7), we derive an approximate expression for the lasing intensity

$$|a_1|^2 \approx \frac{\Delta d}{d_{\text{thresh}}} \frac{\int_{\Omega} \epsilon_c(\mathbf{r}) |\mathbf{e}_1(\mathbf{r})|^2 d\Omega \int_{\Omega} D_0(\mathbf{r}) |\mathbf{e}_1(\mathbf{r})|^2 d\Omega}{e_c^{-2} \int_{\Omega} D_0(\mathbf{r}) |\mathbf{e}_1(\mathbf{r})|^4 d\Omega}, \quad (10)$$

where we have assumed that we are operating in the weak pumping limit (just above threshold), where $d \approx d_{\text{thresh}}$ and where $\Delta d = d - d_{\text{thresh}} \ll 1$ represents the pumping strength above threshold. For more details on this derivation, consult Appendix B. Note that evaluating Equations (9)–(10) still requires one to solve a linear eigenproblem to determine the modes of the passive cavity, as in SPA-SALT. Nevertheless, assuming a high- Q cavity ($Q \gtrsim 100$), we can eliminate this requirement by exciting the field in the cavity with an external source (e.g., a waveguide coupled to the cavity), since the field in the cavity will be approximately proportional to the passive cavity mode $\mathbf{e}_1$ [3, 45]. Although this eliminates the requirement of an eigensolve, it does not include information about the efficiency of coupling between the cavity mode and the output channel. In the next section, we will use TCMT to encode the out-coupling into a single, convenient FOM.

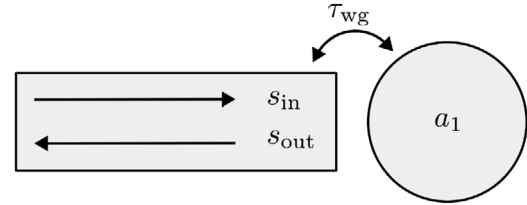


FIGURE 2 | Coupled optical system composed of a waveguide guiding an input field with amplitude s_{in} and an output field with amplitude s_{out} , coupled to a cavity with an amplitude a_1 through the decay rate τ_{wg} .

2.3 | Nonlinear FOM via Temporal Coupled Mode Theory

To define a FOM that can be evaluated by solving a single linear system of equations and includes information about the coupling into the output channel, we use the expression for the amplitude in Equation (10) in the context of TCMT. For a system with a sufficiently strong resonance (e.g., $Q \gtrsim 100$ [3]) coupled to a discrete set of input/output ports, TCMT derives a minimal set of equations coupling the resonant amplitude with the amplitudes of the port modes, constrained by energy conservation and reciprocity to be determined purely in terms of the coupling rates $1/\tau$ (lifetimes τ) [3]. Such a coupled optical system is illustrated in Figure 2, where a cavity mode with amplitude a_1 (normalized so that $|a_1|^2$ is the energy in the cavity), *total* decay rate of the resonance $1/\tau_1$, and quality factor $Q = \omega_1 \tau_1/2$, is coupled to a waveguide with an input amplitude s_{in} and output amplitude s_{out} (normalized so that $|s|^2$ is power) through a decay rate $1/\tau_{\text{wg}}$. Note that the *total* decay rate $1/\tau_1$ includes the decay rate $1/\tau_{\text{wg}}$ into the waveguide plus the decay rates into other loss channels such as absorption and/or radiation [3].

2.3.1 | Emission Problem

In a direct formulation of the emission problem, the emitters in the gain medium will generate energy in the cavity mode, which then couples into the output waveguide mode. Using the TCMT equations, one can calculate the output power when there is no input waveguide mode ($s_{\text{in}} = 0$), as [3]

$$P_{\text{out}} = |s_{\text{out}}|^2 = \frac{\omega_1}{Q_{\text{wg}}} |a_1|^2, \quad (11)$$

where $Q_{\text{wg}} = \omega_1 \tau_{\text{wg}}/2$ is the quality factor for the coupling between cavity and waveguide. By substituting the expression for the amplitude in Equation (10) and the expression for the threshold in Equation (9), the TCMT output power becomes

$$P_{\text{out}} \approx \Delta d \omega_1 \frac{Q}{Q_{\text{wg}}} \frac{\left(\int_{\Omega} D_0(\mathbf{r}) |\mathbf{e}_1(\mathbf{r})|^2 d\Omega \right)^2}{e_c^{-2} \int_{\Omega} D_0(\mathbf{r}) |\mathbf{e}_1(\mathbf{r})|^4 d\Omega}. \quad (12)$$

Note that, to evaluate this expression, one still needs to solve a single linear eigenvalue problem to determine the quality factors and mode profiles.

2.3.2 | Reciprocal Problem

Instead, we now consider the reciprocal problem, where the input port excites the resonance (often a useful approach for ensembles of emitters [16]). In this case, the TCMT equations imply a simple linear relationship between the input power and the resonant amplitude [3]

$$U = |a_1|^2 = \frac{2\tau_1^2}{\tau_{\text{wg}}} |s_{\text{in}}|^2 = \frac{1}{\omega_1} \frac{Q^2}{Q_{\text{wg}}} P_{\text{in}}. \quad (13)$$

In the high- Q system, the field from the reciprocal solve $\mathbf{E}_r$ is almost exactly equal to the cavity mode plus an $\mathcal{O}(1/\sqrt{Q})$ error [3]; thus, the energy $|a_1|^2$ can also be readily calculated by evaluating an integral of the reciprocal electric field in the cavity: $U = \int_{\Omega} \epsilon_c(\mathbf{r}) |\mathbf{E}_r(\mathbf{r})|^2 d\Omega$. Similarly, with this argument, the reciprocal field $\mathbf{E}_r$ can also be used as a substitute for the modal field in Equations (9)–(12).

2.3.3 | Nonlinear FOM for Nanolaser Optimization

By solving the reciprocal problem in the weak pumping regime ($\Delta d \ll 1$), where we can model the pump power as being proportional to the pump strength ($P_{\text{pump}} \propto d \approx d_{\text{thresh}}$), we can define a nonlinear lasing FOM

$$\frac{P_{\text{out}}}{P_{\text{pump}}} \propto \frac{\left(\int_{\Omega} D_0(\mathbf{r}) |\mathbf{E}_r(\mathbf{r})|^2 d\Omega \right)^3}{\int_{\Omega} D_0(\mathbf{r}) |\mathbf{E}_r(\mathbf{r})|^4 d\Omega} = \text{FOM}. \quad (14)$$

where we have used the expression of the output power in Equation (12), and the lasing threshold expression in Equation (9). We have left out proportionality factors and have substituted the energy integral with the expression derived in Equation (13). Although the “efficiency” ($P_{\text{out}}/P_{\text{pump}}$) on the left-hand-side of Equation (14) is ≤ 1 , eliminating the proportionality factors results in the FOM defined by the right-hand-side taking arbitrary (nonnegative) values. Notice that all quality factors simplify, and we are left with an expression that only depends on reciprocal field integrals, meaning that the FOM can be evaluated with a single reciprocal solve (excited from the output waveguide). Moreover, this FOM $\sim |\mathbf{E}_r|^6/|\mathbf{E}_r|^4 \sim |\mathbf{E}_r|^2$ is roughly proportional to the energy in the cavity ($\sim |\mathbf{E}_r|^2$) and thus to the Q factor. This fact will ensure that maximizing the FOM yields high- Q optimized cavities, making the perturbative single-pole approximation more and more accurate as the optimization progresses (numerical validation of this high- Q approximation is provided in Appendix D for optimized nanolasers).

2.3.4 | LDOS and the Single-Emitter Limit

To show how the nonlinear/first-principles FOM compares to more conventional cavity-optimization approaches, we introduce a “naive” generalization of the LDOS, which *heuristically* modifies the definition of the LDOS to account for out-coupling efficiency (through a reciprocal solve $\mathbf{E}_r$) and a distributed gain

medium D_0

$$\text{FOM}_{\text{naive}} = \int_{\Omega} D_0(\mathbf{r}) |\mathbf{E}_r(\mathbf{r})|^2 d\Omega. \quad (15)$$

This FOM targets the overlap of the electric-field intensity of the reciprocal field with the gain distribution. Like our nonlinear FOM, this naive FOM is also proportional to the electric-field intensity in the cavity ($\sim |\mathbf{E}_r|^2$) and thus to the Q factor (all other things being equal, though this FOM is also affected by the coupling/extraction efficiency), which will also tend to yield high- Q optimized cavities, with strong field in the gain medium. Similarly, it will favor a low lasing threshold via Equation (9). Indeed, we observe these properties in the inverse-design results of Section 3.

Interestingly, in the single-emitter limit, where the gain distribution is modeled as being concentrated at a single point $\mathbf{r}'$, i.e., $D_0(\mathbf{r}) \propto \delta(\mathbf{r} - \mathbf{r}')$, the nonlinear FOM (Equation (14)) and the naive FOM (Equation (15)) become *equivalent*, reducing to an LDOS-like FOM when solving the reciprocal problem [46]: $\text{FOM} \propto |\mathbf{E}_r(\mathbf{r}')|^2$. Whereas the usual definition of LDOS would correspond to the *total* power expended by an orientation-averaged dipole source at $\mathbf{r}'$ [6], this FOM corresponds (via reciprocity [46, App. C]) to the dipole power emitted only into the output channel, a form of extraction efficiency. Such a point-gain FOM is therefore reminiscent of earlier efforts [21–25] to optimize cavities by maximizing LDOS or extraction efficiency.

However, for extended active regions with a gain region whose diameter is not small compared to the wavelength ($\gtrsim \lambda$), the naive and nonlinear objective functions are inequivalent, and we will show in Section 3.1 that the resulting geometries and performance differ substantially, highlighting the importance of deriving the correct first-principles FOM rather than using heuristic targets.

Another interesting situation is that of a single point-like gain region, such as a quantum dot, whose location randomly varies from one fabrication to the next. In that case, one might wish to target the *average* performance (the expected value $E[\text{FOM}]$). Since the point-like gain FOM is $\propto |\mathbf{E}_r(\mathbf{r}')|^2$, performing an ensemble average over gain locations $\mathbf{r}'$ with a probability distribution $\mathcal{P}(\mathbf{r}')$ immediately leads to $E[\text{FOM}] \propto \int \mathcal{P}(\mathbf{r}) |\mathbf{E}_r(\mathbf{r})|^2 d\Omega$, identical to Equation (15) except that D_0 is replaced by $\mathcal{P}$.

2.4 | Accounting for Steady-State Diffusion in the Gain Medium

In semiconductor lasers with extended media, i.e., as opposed to isolated emitters such as quantum dots, it is paramount to account for diffusion of excited carriers. Thus, while the spatial profile of the lasing mode leads to a spatial variation of the local rate of stimulated emission and hence a spatially varying depletion of excited states, gain diffusion smoothens out the effective gain profile. As introduced in the complex-SALT (C-SALT) model [4], this effect can be accounted for by modeling a bulk semiconductor gain medium in the free-carrier approximation, where carrier–carrier Coulomb interactions are neglected. In this approximation, one can include diffusion in

Equation (7) by applying a set of linear operators

$$\Delta \varepsilon_{\mathbb{S}}(\mathbf{r}) = (\mathbb{S}^{-1} + \mathbb{I} e_c^{-2} |\mathbf{E}_L|^2)^{-1} [D_0](\mathbf{r}) d \quad (16)$$

where $\mathbb{I}$ is the identity operator, and the exponentially damped diffusion is encoded in the operator $\mathbb{S}^{-1} = \mathbb{I} + \nabla \cdot (R_V^2(\mathbf{r}) \nabla)$, where $R_V(\mathbf{r})$ is a diffusion lengthscale determined by the spatially- and design-dependent diffusion coefficient combined with the damping/recombination rate. In this notation, an operator is applied as $\mathbb{I}[b]$, where $\mathbb{I}$ is the operator and b is a scalar field. As an example, computing $u = \mathbb{S}[b]$ corresponds to solving for the scalar field u in the diffusion problem $\mathbb{S}^{-1}u = b$. Note that for small diffusion $R_V \ll \lambda$ we recover the original expression in Equation (5) because $\mathbb{S} \approx \mathbb{I}$.

Similar to the derivations in Section 2.2, we calculate the pumping strength needed to reach the lasing threshold by employing perturbation theory for small $|\mathbf{E}_L|^2 \sim |a_1|^2$

$$d_{\text{thresh}} = \frac{1}{Q} \frac{\int_{\Omega} \varepsilon_c(\mathbf{r}) |\mathbf{e}_1(\mathbf{r})|^2 d\Omega}{\int_{\Omega} \mathbb{S}[D_0](\mathbf{r}) |\mathbf{e}_1(\mathbf{r})|^2 d\Omega}. \quad (17)$$

In this case, the lasing threshold can be decreased by overlapping the mode with the diffused gain profile $\mathbb{S}[D_0]$ in the cavity.

In a similar fashion to the derivations in Section 2.3, we use TCMT to calculate the nonlinear lasing FOM that accounts for diffusion

$$\text{FOM} \propto \frac{(\int_{\Omega} \mathbb{S}[D_0](\mathbf{r}) |\mathbf{E}_r(\mathbf{r})|^2 d\Omega)^3}{\int_{\Omega} \mathbb{S}[|\mathbf{E}_r|^2 \mathbb{S}[D_0]](\mathbf{r}) |\mathbf{E}_r(\mathbf{r})|^2 d\Omega} \quad (18)$$

where the diffusion is applied to both the gain profile in the numerator, and to the product of the diffused gain profile $\mathbb{S}[D_0]$ and the electric-field intensity of the reciprocal field in the denominator. This highlights the fact that diffusion will not only affect the gain medium but will also diffuse depletion of the emitters in the gain region due to the hole-burning effect. For a more detailed derivation of the expressions in this section, refer to Appendix C.

2.5 | Topology Optimization Formulation

In this section, we review the standard density-based TopOpt formulation [27, 47] that we employ to describe and optimize our degrees of freedom, which is mostly independent of the choice of FOM. Density-based TopOpt parameterizes the distribution of materials in space by introducing a design field, the “density” $\rho(\mathbf{r}) \in [0, 1]$, where the extremes $\rho = 0$ and $\rho = 1$ correspond to the two materials, while unphysical intermediate materials are also permitted temporarily as the design evolves. To regularize the optimization problem by introducing a weak sense of geometric lengthscale (ensuring convergence as the computational resolution is increased), one applies a filtering and thresholding scheme to the design variables. The thresholding, which pushes the design toward a “binarized” structure in which intermediate material are excluded, is performed via a smoothed Heaviside

projection [48]

$$\hat{\rho} \equiv \Theta_{\beta, \eta}(\tilde{\rho}) = \frac{\tanh(\beta \cdot \eta) + \tanh(\beta \cdot (\tilde{\rho} - \eta))}{\tanh(\beta \cdot \eta) + \tanh(\beta \cdot (1 - \eta))}, \quad (19)$$

where $\tilde{\rho}$ is the physical field, $\Theta_{\beta, \eta}(\tilde{\rho})$ is the thresholding function, $\beta \in [1, \infty)$ and $\eta \in [0, 1]$ are hyper-parameters that control the threshold sharpness and value respectively. The filtered design field $\tilde{\rho}$, which regularizes the problem by introducing a minimum lengthscale r_f , is obtained using a Helmholtz-based filter [49]

$$-\left(\frac{r_f}{2\sqrt{3}}\right)^2 \nabla^2 \tilde{\rho} + \tilde{\rho} = \rho, \quad (20)$$

where r_f is a filter radius. The projection steepness β is gradually increased as the optimization progresses: a small β allows grayscale structures in which the topology can change smoothly, while a larger β binarizes the structure but slows convergence (so the largest β is only imposed at the end, when the structure is nearly converged) [50, 51]; the precise optimization details are given in Appendix E.

The filtering step regularizes the problem, but does not strictly exclude features that have arbitrarily small lengthscales. In order to exclude such non-manufacturable results, we employ a standard technique to impose additional geometric minimum-lengthscale constraints [52, 53] with a minimum feature size just above the single-pixel level ($\gtrsim 40$ nm) on both the solid and void regions, with the details in Appendix E.

The projected density $\hat{\rho}$ is mapped onto a physical material (and gain) by a straightforward interpolation. In particular, the refractive index n is given by a linear interpolation

$$n = n_0 + \hat{\rho}(n_1 - n_0), \quad (21)$$

where subscript 1 denotes the material and 0 the background. We also model a design-dependent extinction coefficient

$$\kappa = -(\alpha + \alpha' \hat{\rho}(1 - \hat{\rho})), \quad (22)$$

where α is a coefficient that introduces artificial optical losses to broaden the frequency response of high-Q optical resonances, thus facilitating the optimization problem [22]; and α' is an attenuation factor that discourages intermediate values of $\hat{\rho}$ by introducing optical losses. As the optimization progresses, α is gradually “turned off” to zero (eliminating the unphysical damping) while α' is gradually “turned on” (penalizing unphysical intermediate materials). The gain medium is also dependent on $\hat{\rho}$, because we assume that gain is only present in the solid material. Furthermore, in Section 3, we explore gain regions of varying diameters, described by a Gaussian profile around the center $\mathbf{r}_0$ of the cavity, yielding the equation

$$D_0(\mathbf{r}, \hat{\rho}) = \varepsilon_{c(\mathbf{r})} \hat{\rho}(\mathbf{r}) e^{-|\mathbf{r} - \mathbf{r}_0|^2 / 2\sigma_g^2}, \quad (23)$$

where σ_g parametrizes the width of the Gaussian. Future work may replace this ad-hoc gain profile with an explicit physical model of a pumping process, such as optical absorption, as discussed in Section 4.

3 | Computational Results

In this section, we apply the theoretical framework reviewed and developed in the previous sections to optimize nanocavity-based laser devices. First, we compare our developed first-principles FOM with a heuristic generalization of the LDOS (Section 2.3.4) through the inverse design 2d cavities (Section 3.1), demonstrating the increasing importance of a first-principles model as the diameter of the gain region increases. Then, we study how steady-state diffusion effects influence the optimized nanocavity designs (Section 3.2). In all cases, the permittivity of the solid regions is $\epsilon = 12$ (characteristic of semiconductors) and of the void regions is $\epsilon = 1$ (air). The gain medium material is modeled as a III–V semiconductor-like material with a direct bandgap. Lastly (Section 3.3), we consider the inverse design of more realistic 3d cavities in semiconductor-on-insulator system (SOI, $\epsilon = 1.44^2$ for the substrate). (Detailed simulation and optimization parameters are given in Appendix E.)

In all cases, the laser is optimized for a wavelength of 1550 nm, and we impose a minimum lengthscale [52] (Section 2.5) just above the single-pixel level ($\gtrsim 40$ nm) on both solid and void regions. The design region is $2.325\ \mu\text{m} \times 2.325\ \mu\text{m}$, connected to a single-mode output waveguide (width 500 nm), with PML absorbing boundary layers [54, 55] along with absorbing boundary conditions [56]. The 2d simulations employ the H_z (out-of-plane $\mathbf{H}$, in-plane $\mathbf{E}$) polarization, whereas the 3d simulations use the analogous mostly- H_z waveguide mode. All simulations employ a finite-element discretization whose details are given in Appendix E, with resolution ≈ 40 nm in the design region. Antisymmetric mirror-symmetry boundary conditions (Neumann for H_z , equivalent to a perfect electric conductor) are imposed to halve the computational region (bisecting the waveguide). Although such a non-convex optimization problem may, in principle, have non-symmetric local optima (and TopOpt often exhibits many local optima with similar performance [57, 58]), we found that symmetric optima were often found even if we did not impose symmetry, and imposing mirror symmetry reduces computational costs. Except where otherwise noted, all optimizations began with a homogeneous $\rho = 0.5$ starting structure.

3.1 | 2D: Nonlinear vs. Naive FOM

To showcase the difference of optimizing for a naive generalization of the LDOS (Equation (15), Section 2.3.4) or for the first-principles nonlinear FOM (Equation (14), Section 2.3.3), in Figure 3 we compare the performance of optimized nanolasers for different gain region sizes: $\sigma_g \in \{25\ \text{nm}, 50\ \text{nm}, 100\ \text{nm}, 250\ \text{nm}, 400\ \text{nm}, 500\ \text{nm}\}$. Half of the cavities are optimized for the naive FOM (green), while the other half are optimized for the derived nonlinear FOM (yellow). To make the FOM easier to interpret, reproduce, and compare across nanolaser platforms, we normalize the nonlinear FOM by dividing by $\zeta(\sigma_g) = \max\{|\mathbf{E}_{\text{in}}|^2\} \left(\int_{\Omega} D_0^*(\sigma_g) d\Omega\right)^2$, where $\mathbf{E}_{\text{in}}$ is the input-waveguide electric field and $D_0^* = e^{-|\mathbf{r}-\mathbf{r}_0|^2/2\sigma_g^2}$ is the Gaussian gain profile with the structure (ϵ, ρ) dependence removed. Note that the results have been cross checked to ensure that indeed a design optimized for a certain domain size is the best for that domain size.

As pointed out in Section 2.3.4, in the single-emitter limit where the active region is point-like, the naive and nonlinear FOM become equivalent, and in consequence we observe in Figure 3b that the optimized cavities become nearly equivalent for small $\sigma_g \ll \lambda$. In this single-emitter limit, the cavity geometry exhibits a central bowtie-like feature—a structure commonly arising in LDOS optimization [21–25]—which is known to enhance light-matter interaction at a single point [31, 59, 60] by concentrating the in-plane electric field at the emitter location thanks to a field singularity [10, 61] at any sharp tip. (Here, the bowtie sharpness is limited by our $\gtrsim 40$ nm lengthscale constraints.) Due to the strong field concentration at the cavity-tip, the normalized FOM/ $\zeta(\sigma_g)$, has a larger value compared to the cavities that were optimized for larger gain-region sizes.

As the active region size increases, the system transitions beyond the single-emitter regime, and the cavity designs optimized for the naive and nonlinear FOMs begin to differ significantly as the gain diameter approaches the wavelength ($\sigma_g \gtrsim \lambda/n_1 \sim 450$ nm). In both cases, the spatial averaging of the emission means that arbitrarily sharp tips are no longer beneficial (because $\int |\mathbf{E}|^2$ is finite) [31]. For our largest gain diameter ($\sigma_g = 500$ nm, Figure 3d), the structure optimized using the nonlinear FOM achieves a performance $\approx 3\times$ greater than the structure optimized with the naive FOM. We attribute this to the nonlinear “hole-burning” effect (Section 2.1) captured in the denominator of the nonlinear FOM (Equation (14)) but not present in the naive FOM (Equation (15)), which penalizes strong spatial field localization. As illustrated in Figure 4, the electric-field distribution shows that the cavity optimized for the naive FOM concentrates the field at its center, whereas the cavity optimized for the nonlinear FOM distributes the field more homogeneously around the active region D_0 , reducing hole burning and leading to a higher nonlinear FOM. Moreover, as the active region size becomes larger, the normalized FOM (FOM/ $\zeta(\sigma_g)$) decreases (Figure 3a), and transitions from the sharp-tip large-enhancement regime to distributing the electric field within the active material, which results in the normalized FOM becoming roughly proportional to Q for larger active region sizes (Section 2.3.3). Note that the non-normalized FOM increases by several orders of magnitude as the active-region volume grows, simply reflecting the larger gain medium emitting light into the lasing mode. We emphasize that field singularities at sharp tips play a negligible role in most of the designs. As discussed above, they are only favored when the gain region is effectively point-like and the two FOMs become LDOS-like, whereas they are explicitly penalized in the nonlinear FOM for an active region with size comparable to the wavelength of light. This suppression of singular features contributes to the intrinsic robustness of the optimized cavities, as the performance does not rely on the minimum lengthscale of the geometry.

Finally, we validate a posteriori one of the main assumptions in the first-principles modeling based on the SPA-SALT (Section 2.2): that the optimized cavities are indeed high- Q cavities. By means of a post-processing eigensolve on the optimized cavities, we find that the Q s of the lasing modes are in the range $Q \in [350, 1000] \gtrsim 100$, which validates the SPA-SALT and TCMT approximations [3, 43] as shown in Appendix D. Not only that, as pointed out in Section 1, there is a trade-off between designing a high- Q cavity and a cavity that couples efficiently to an output port. By computing the power fraction in the waveguide relative

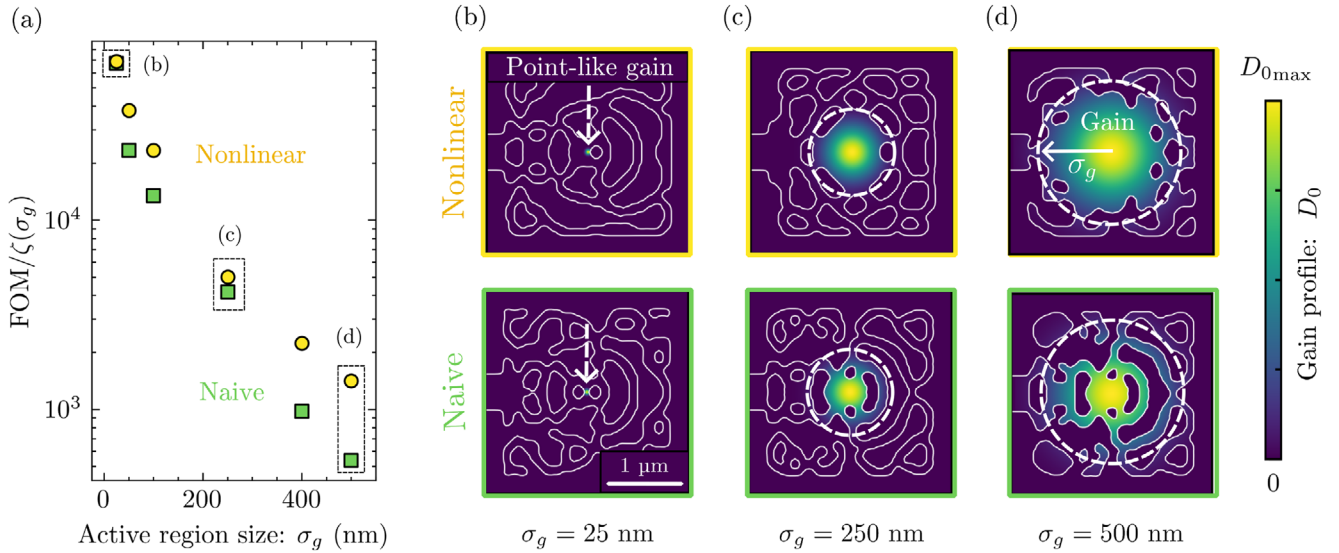


FIGURE 3 | Performance of inverse designed devices for increasing active region sizes σ_g . (a) Evaluation of the nonlinear FOM normalized by the gain region size-dependent factor $\zeta(\sigma_g) = \max \{ |\mathbf{E}_{\text{in}}|^2 \} \left(\int_{\Omega} D_0^*(\sigma_g) d\Omega \right)^2$, for designs optimized for the naive FOM (green) and nonlinear FOM (yellow). The gain profile D_0 is shown in (b), (c), and (d) in max-normalized units for increasing active region sizes. The standard deviation σ_g of the gain profile is marked in a dashed white line. In the gain profile for the nonlinear device in (d), the active region size σ_g is explicitly shown for illustration purposes.

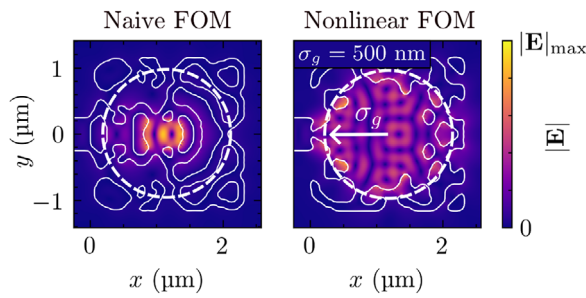


FIGURE 4 | Electric-field norm $|\mathbf{E}(\mathbf{r})|$ in devices optimized for a gain medium of standard deviation $\sigma_g = 500$ nm. The fields are expressed in max-normalized units.

to the total lasing-mode power in the post-processing eigensolve, we evaluate the lasing mode's optical extraction efficiency in the topology-optimized cavities. As expected from the definition of the nonlinear FOM (Equation (14)), which directly incorporates out-coupling effects, we find values of the optical extraction efficiency ≥ 0.9 for most of the designs, highlighting the ability of the FOM to yield high-Q cavities with high optical extraction efficiencies. We finally note that we have also assessed fabrication robustness by evaluating how systematic over- and under-etching would affect performance using the nonlinear FOM for the device with $\sigma = 500$ nm. Following [62], a change in threshold value $\Delta\eta = 0.1$ corresponds to a dimensional variation of $\Delta l \approx 10$ nm, yielding a FOM scaled by a factor $\approx \times 0.20$ when evaluated at the new (shifted) resonance frequency as determined by an eigensolver. Moreover, to study how realistic random roughness (e.g., surface roughness) affects the design, we have considered a random variation of $\Delta\eta = 0.1$ with a spatial correlation of 10 nm, and see that the FOM is scaled by a factor $\approx \times 0.55$ for such perturbations. This provides a quantitative estimate of expected performance tolerance to nanoscale variations for the optimized high-Q nanolaser designs.

3.2 | 2D: Gain Diffusion

Gain diffusion is important in extended gain media, such as bulk materials and quantum wells, when the extent of the gain region becomes comparable to or larger than the diffusion length. In such cases, the diffusion counteracts the effect of spatial-hole burning. To understand the impact of gain diffusion on the optimization results we use a diffusion model based on the C-SALT framework [4](Section 2.4) to analyze the performance of the cavity previously optimized for a gain medium of size $\sigma_g = 250$ nm (Figure 3), where diffusion was neglected. Using this design as an initial guess, we re-optimize the cavity in Figure 5 using the diffusion FOM (Equation (18)), assuming a diffusion length of $R_0 = 5 \mu\text{m}$ comparable to typical electron and hole diffusion lengths in semiconductor materials such as Si [63, 64], InP [64], (In,Ga)N [65], InAs [64], GaAs [64], and InGaAs [66]. Note that as the diffusion lengths decrease, the system approaches the non-diffusive limit ($R_0^2 \ll 1$, Equation (16)), and the optimization results will converge toward those presented in Section 3.1.

Evaluating the diffusion-corrected performance of the device optimized without considering diffusion (FOM/ $\zeta \approx 1119$, “Nonlinear FOM” in Figure 5) reveals a significant 2-fold drop in FOM compared to the diffusion-optimized structure (FOM/ $\zeta \approx 2118$, “Diffusion FOM” in Figure 5), underscoring the importance of including diffusion effects in the optimization process. This reduction in FOM can be attributed to the diffusion of carriers within the active region, as shown in the diffused gain profile $S[D_0]$, which reduces the field overlap in a structure designed for the nominal Gaussian gain profile. However, by optimizing for the FOM that incorporates diffusion, the cavity design is able to compensate for this effect: both changing the field profile to match the modified gain distribution, and by changing the structure to control the diffusion itself. In particular, as shown in Figure 5, the optimizer disconnects the cavity from both the waveguide and low-field regions, while

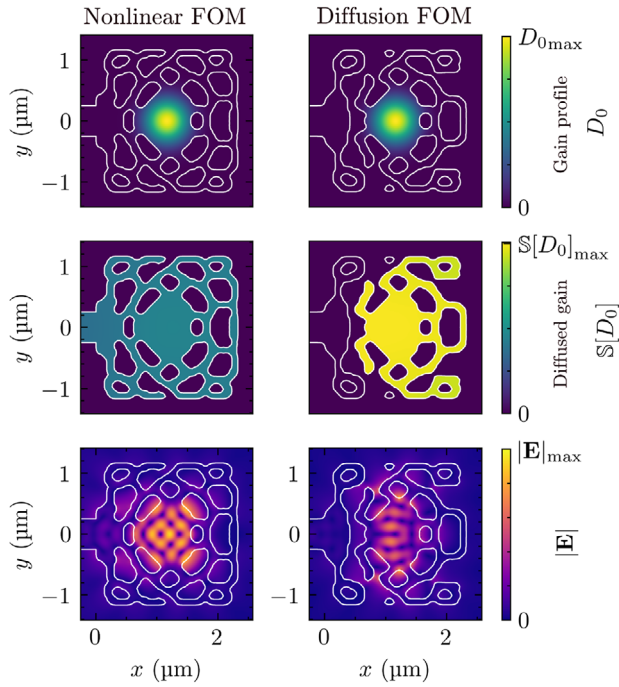


FIGURE 5 | Spatial profile of the gain medium $D_0(\mathbf{r})$ with a standard deviation $\sigma_g = 250$ nm, the diffused gain medium $S[D_0](\mathbf{r})$, for a diffusion length of $R_0 = 5 \mu\text{m}$, and the electric-field norm $|\mathbf{E}(\mathbf{r})|$ of the cavity mode for the inverse designed cavity not accounting (left) and accounting (right) for diffusion. All fields are expressed in max-normalized units.

also removing material from areas with weak electric field. This confines the emitters or carriers to the high-field cavity region and improves the overlap between the diffused gain region $S[D_0]$ and the electric field distribution (which is also more spread out than in the non-diffusion design).

3.3 | Cavity Design of a 3D Physics Model

In this section, we demonstrate that our approach is computationally efficient enough to attack fully 3D systems, since it is no more expensive than traditional linear inverse design. In particular, we consider an example nanolaser cavity designed for typical semiconductor-on-insulator platforms, in which a 220 nm semiconductor slab is patterned on top of a silica ($n = 1.44$) substrate, with air above. The design region is otherwise similar to the 2d geometries, consisting of a 2d pattern and waveguide extruded with constant cross sections in the semiconductor layer, where voids in this layer are filled with air and the active material takes up the entire thickness (extrusion length) of the slab and is embedded within the semiconductor structure. PML+absorbing boundaries are used above and below the slab, separated from the semiconductor layer by 720 nm. Additional implementation details are provided in Appendix E.

Using this 3d model, we again optimize nanolaser cavities with a gain medium size of $\sigma_g = 500$ nm for the naive and nonlinear FOMs (Figure 6), with no gain diffusion. Similar to the 2d case, the device optimized for the naive FOM creates a strong field concentration close to the cavity center, while the device optimized for the nonlinear FOM spreads the field

more evenly over the active material region to counteract hole-burning effects. However, and in contrast to the 2d optimization results (Section 3.1), in 3d there is a more modest increase of the FOM when targeting nonlinear (FOM/ $\zeta \approx 148$) vs the naive FOM (FOM/ $\zeta \approx 90$)—there is an increase of a factor of ≈ 1.6 (instead of the factor of 3 in Section 3.1). We attribute this result qualitatively to the fact that localizing a high-Q resonance is more difficult in a 3d slab system than in 2d, because of the need to minimize out-of-plane radiation. That is, in 3d, more design freedom is devoted to increasing Q than in 2d, leaving less design freedom to optimize the field profile for hole-burning effects. Nevertheless, the increase in the FOM is still a substantial improvement, with no additional computational cost over a naive approach. Post-processing the 3d results through an eigensolve again validates the high-Q assumption for SPA-SALT and TCMT analysis, with values of $Q \approx 150$, while the optical extraction efficiency is moderately high ≥ 0.75 , lower than in 2d (Section 3.1) presumably due to out-of-plane radiation for this tiny ($\approx 1.5\lambda \times 1.5\lambda$) cavity. We expect that higher Q and efficiency should occur for larger design regions, similar to previous nanocavity designs.

4 | Conclusion

In this paper, we have shown that by careful analysis of a lasing model, exploiting perturbative analysis valid in high-Q cavities, we can obtain an optimization figure of merit (FOM) for the input-to-output efficiency of the laser that incorporates a wide range of important phenomena—such as resonant enhancement, spatial hole-burning, and gain diffusion—with little or no additional computational cost compared to simpler FOMs such as the LDOS. Only a single *linear* “reciprocal” Maxwell solve is required to evaluate our FOM, plus two scalar damped-diffusion solves if gain diffusion is included. The results, in both two and three dimensions, show substantial benefits to such a first-principles approach compared to heuristic field-intensity/LDOS-like figures of merit, especially for a finite-diameter gain region.

While we believe that our new FOM should be immediately valuable to inverse design of nanolasers, it also offers a starting point for many future refinements. An obvious next step would be to replace our ad-hoc gain profile D_0 with a full first-principles model of the pumping process itself, since that process is likely to also depend somewhat on the geometry. For example, optical pumping can be modeled as a second Maxwell solve at the pump wavelength (e.g., for vertical illumination), with D_0 being computed from the pump absorption profile [4, 37, 67]. This is conceptually similar to the coupled pump–emission solves that have been applied to inverse design of Raman scattering [29–31], scintillation [32], and nonlinear harmonic generation [33, 34]. Moreover, one could include a geometry-dependent effect on the pump power to reach transparency (Section 2.1), as well as other effects such as absorption in the non-pumped region. Alternatively, modeling electrical pumping would involve solving a carrier-transport equation for D_0 [68, 69], again coupled to the reciprocal lasing solve. Also, for the ultra-small active regions considered in nanolasers, pump blocking due to bandfilling effects become an important effect [11]. Even more complicated pumping models might involve nonlinear interactions between the pumping and emission processes, but since the emission should still simplify via a resonance approximation, we are hope-

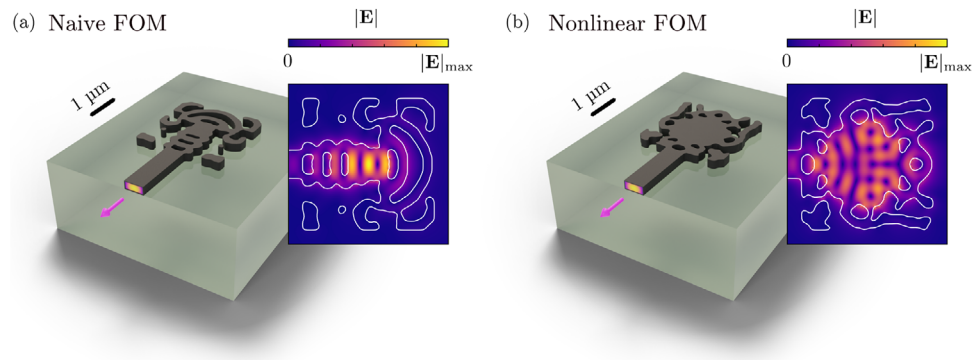


FIGURE 6 | Inverse designed devices in three dimensions. The devices lase into the cavity mode (inset plot) and out-couple to the fundamental TM mode of the output waveguide (pink arrow). The electric-field norm is given in max-normalized units. (a) Device optimized for the naive FOM. (b) Device optimized for the nonlinear FOM.

ful that such effects will still be a tractable nonlinear solve (e.g., by a few Newton steps [40] for pre-selected poles), which would be differentiable by implicit adjoint methods [41]. In principle, optimization could also be applied to the full SALT equations, with the perturbative approach providing an initial guess that may reduce the difficulty of pole tracking—such a full-SALT optimization may become necessary if one wants to drive the system into the multi-mode regime. (Even without optimization, a SALT solve could be useful to quantify the accuracy of the SPA-SALT and TCMT approximations, although based on previous work and the numerical validation provided in Appendix D, we expect the errors in such perturbative methods to be small for high- Q single-mode cavities like the results in this paper.) Another effect that might be interesting to include in an optimization FOM is the lasing linewidth, for which accurate first-principles formulas in terms of the lasing mode have recently been derived that include the full effect of nano-patterned geometries [38], and which should be equally applicable to our reciprocal-solve estimate of the lasing mode in the high- Q regime. It has already been experimentally demonstrated that the distribution of active material in a nanostructured cavity can be used to significantly decrease the linewidth of a nanolaser [70].

As we commented in Section 2.3.4, the “naive” FOM (15) that we employ for comparison purposes is actually a *correct* FOM for situations in which a *single* emitter is randomly distributed with probability density $\sim D_0$. This situation arises experimentally in systems employing quantum dots [71, 72], color centers [73–75], and similar isolated point-like emitters, whose placement is difficult to precisely control relative to the fabricated pattern. Despite its simplicity, Equation (15) still encapsulates both the perturbative lasing physics and the out-coupling efficiency for such applications.

Acknowledgements

We thank A.D. Stone for useful discussions. We gratefully acknowledge financial support from the Danish National Research Foundation through the NanoPhoton Center for Nanophotonics, grant number DNRF147. A.C. acknowledges support from the Laboratory Directed Research and Development program at Sandia National Laboratories. This work was performed in part at the Center for Integrated Nanotechnologies, an

Office of Science User Facility operated for the U.S. Department of Energy (DOE) Office of Science. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. DOE’s National Nuclear Security Administration under Contract No. DE-NA-0003525. S.G.J. was supported in part by the Simons Foundation collaboration on Extreme Wave Phenomena, and by the U.S. Army Research Office (ARO) through the Institute for Soldier Nanotechnologies (ISN) under award no. W911NF-23-2-0121. The views expressed in the article do not necessarily represent the views of the U.S. DOE or the United States Government.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. H. E. Türeci, A. D. Stone, and B. Collier, “Self-Consistent Multimode Lasing Theory for Complex or Random Lasing Media,” *Physical Review A* 74 (Oct. 2006): 043822, <https://doi.org/10.1103/PhysRevA.74.043822>.
2. L. Ge, Y. D. Chong, and A. D. Stone, “Steady-State Ab Initio Laser Theory: Generalizations and Analytic Results,” *Physical Review A* 82, no. 6 (Dec. 2010), <http://doi.org/10.1103/PhysRevA.82.063824>.
3. J. D. Joannopoulos, S. G. Johnson, J. N. Winn, and R. D. Meade, *Photonic Crystals: Molding the Flow of Light*, 2nd ed (USA: Princeton University Press, 2008).
4. A. Cerjan, Y. D. Chong, and A. D. Stone, “Steady-State Ab Initio Laser Theory for Complex Gain Media,” *Optics Express* 23, no. 5 (Mar. 2015): 6455–6477, <https://opg.optica.org/oe/abstract.cfm?URI=oe-23-5-6455>.
5. M. Xiong, Y. Yu, Y. Berdnikov, et al., “A Nanolaser with Extreme Dielectric Confinement,” *Science Advances* 11, no. 51 (2025): ead3865.
6. A. Oskooi and S. G. Johnson, “Electromagnetic Wave Source Conditions,” in *Advances in FDTD Computational Electrodynamics: Photonics and Nanotechnology*, A. Taflov, A. Oskooi, and S. G. Johnson, Eds (Boston: Artech, 2013), ch. 4 65–100.
7. Y. Yamamoto, S. Machida, and G. Björk, “Microcavity Semiconductor Laser with Enhanced Spontaneous Emission,” *Physical Review A* 44 (Jul. 1991): 657–668, <https://doi.org/10.1103/PhysRevA.44.657>.
8. E. K. Lau, A. Lakhani, R. S. Tucker, and M. C. Wu, “Enhanced Modulation Bandwidth of Nanocavity Light Emitting Devices,” *Optics*

- Express 17, no. 10 (May. 2009): 7790–7799, <https://opg.optica.org/oe/abstract.cfm?URI=oe-17-10-7790>.
9. J. Mørk, M. Xiong, K. Seegert, M. Marchal, G. Dong, E. Dimopoulos, E. Semenova, K. Yvind, and Y. Yu, “Nanostructured Semiconductor Lasers,” *IEEE Journal of Selected Topics in Quantum Electronics* 31, no. 2: Pwr. and Effic. Scaling in Semiconductor Lasers (2025): 1–17.
 10. Y. Ouyang, L. Hongyi, Z. Zhao, W. Mao, and R. Ma, “Singular Dielectric Nanolaser with Atomic-Scale Field Localization,” *Nature* 632 (2024): 1–7.
 11. M. Saldutti, Y. Yu, and J. Mørk, “The Onset of Lasing in Semiconductor Nanolasers,” *Laser & Photonics Reviews* 18, no. 11 (2024): 2300840.
 12. O. Yesilyurt, Z. A. Kudyshev, A. Boltasseva, V. M. Shalae, and A. V. Kildishev, “Efficient Topology-Optimized Couplers for On-Chip Single-Photon Sources,” *ACS Photonics* 8, no. 10 (2021): 3061–3068.
 13. O. T. A. Janssen, A. J. H. Wachtters, and H. P. Urbach, “Efficient Optimization Method for the Light Extraction from Periodically Modulated LEDs Using Reciprocity,” *Optics Express* 18, no. 24 (2010): 24522.
 14. A. Mayer and A. Bay, “Optimization by a Genetic Algorithm of the Light-Extraction Efficiency of a Gan Light-Emitting Diode,” *Journal of Optics* 17 (2015): 025002.
 15. S. Chakravarthi, P. Chao, C. Pederson, et al., “Inverse-Designed Photon Extractors for Optically Addressable Defect Qubits,” *Optica* 7, no. 12 (2020): 1805–1811.
 16. W. Yao, F. Verdugo, R. E. Christiansen, and S. G. Johnson, “Trace Formulation for Photonic Inverse Design with Incoherent Sources,” *Structural and Multidisciplinary Optimization* 65, no. 11 (Nov. 2022), <http://doi.org/10.1007/s00158-022-03389-5>.
 17. H. Chung, “Computational Upper-Limit of Directional Light Emission in Nano-LED via Inverse Design,” *Optics Express* 30, no. 6 (2022): 9008–9020.
 18. S. Kim, J. M. Shin, J. Lee, C. Park, S. Lee, J. Park, D. Seo, S. Park, C. Y. Park, and M. S. Jang, “Inverse Design of Organic Light-Emitting Diode Structure Based on Deep Neural Networks,” *Nanophotonics* 10, no. 18 (2021): 4533–4541.
 19. A. Shultzman, O. Segal, Y. Kurman, C. Roques-Carmes, and I. Kaminer, “Enhanced Imaging Using Inverse Design of Nanophotonic Scintillators,” *Advanced Optical Materials* 11, no. 8 (2023): 2202318.
 20. M. Schubert and A. Hammond, “Fourier Modal Method for Inverse Design of Metasurface-Enhanced Micro-LEDs,” *Optics Express* 31, no. 26 (2023): 42945–42960.
 21. 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,” *Applied Physics Letters* 113, no. 24 (12 2018): 241101, <https://doi.org/10.1063/1.5064468>.
 22. X. Liang and S. G. Johnson, “Formulation for Scalable Optimization of Microcavities via the Frequency-Averaged Local Density of States,” *Optics Express* 21, no. 25 (Dec 2013): 30812–30841, <https://opg.optica.org/oe/abstract.cfm?URI=oe-21-25-30812>.
 23. P. Chao, R. K. Defo, S. Molesky, and A. Rodriguez, “Maximum Electromagnetic Local Density of States via Material Structuring,” *Nanophotonics* 12, no. 3 (2023): 549–557, <https://doi.org/10.1515/nanoph-2022-0600>.
 24. M. Albrechtsen, B. Vosoughi Lahijani, R. E. Christiansen, V. T. H. Nguyen, L. N. Casses, S. E. Hansen, N. Stenger, O. Sigmund, H. Jansen, J. Mørk, and S. Stobbe, “Nanometer-Scale Photon Confinement in Topology-Optimized Dielectric Cavities,” *Nature Communications* 13, no. 1 (Oct. 2022): 6281, <https://www.nature.com/articles/s41467-022-33874-w>.
 25. G. Işiklar, P. T. Kristensen, J. Mørk, O. Sigmund, and R. E. Christiansen, “On the Trade-Off Between Mode Volume and Quality Factor in Dielectric Nanocavities Optimized for Purcell Enhancement,” *Optics Express* 30, no. 26 (Dec. 2022): 47304–47314, <https://opg.optica.org/oe/abstract.cfm?URI=oe-30-26-47304>.
 26. G. H. Ahn, K. Y. Yang, R. Trivedi, et al., “Photonic Inverse Design of On-Chip Microresonators,” *ACS Photonics* 9, no. 6 (2022): 1875–1881.
 27. J. S. Jensen and O. Sigmund, “Topology Optimization for Nano-Photonics,” *Laser & Photonics Reviews* 5, no. 2 (Mar. 2011): 308–321, <https://doi.org/10.1002/lpor.201000014>.
 28. S. Molesky, Z. Lin, A. Y. Piggott, W. Jin, J. Vucković, and A. W. Rodriguez, “Inverse Design in Nanophotonics,” *Nature Photonics* 12, no. 11 (Oct. 2018): 659–670.
 29. R. E. Christiansen, J. Michon, M. Benzaouia, O. Sigmund, and S. G. Johnson, “Inverse Design of Nanoparticles for Enhanced Raman Scattering,” *Optics Express* 28 (2020): 4444–4462.
 30. Y. Pan, R. E. Christiansen, J. Michon, J. Hu, and S. G. Johnson, “Topology Optimization of Surface-Enhanced Raman Scattering Substrates,” *Applied Physics Letters* 119 (2021): 061601.
 31. W. Yao, F. Verdugo, H. O. Everitt, R. E. Christiansen, and S. G. Johnson, “Designing Structures That Maximize Spatially Averaged Surface-Enhanced Raman Spectra,” *Optics Express* 31 (2023): 4964–4977.
 32. C. Roques-Carmes, N. Rivera, A. Ghorashi, et al., “A Framework for Scintillation in Nanophotonics,” *Science* 375, no. 6583 (Feb. 2022), <http://doi.org/10.1126/science.abm9293>.
 33. Z. Lin, X. Liang, M. Lončar, S. G. Johnson, and A. W. Rodriguez, “Cavity-Enhanced Second Harmonic Generation via Nonlinear-Overlap Optimization,” *Optica* 3 (2016): 233–238.
 34. S. A. Mann, H. Goh, and A. Alú, “Inverse Design of Nonlinear Polaronic Metasurfaces for Second Harmonic Generation,” *ACS Photonics* (2023).
 35. H. Haken, *Light: Laser Dynamics*, Vol. 2 (North-Holland Physics Publishing, 1985).
 36. W. E. Lamb, “Theory of an Optical Maser,” *Physical Review* 134 (Jun. 1964): A1429–A1450, <https://doi.org/10.1103/PhysRev.134.A1429>.
 37. A. Cerjan and A. D. Stone, “Steady-State Ab Initio Theory of Lasers with Injected Signals,” *Physical Review A* 90 (2014): 013840, <https://doi.org/10.1103/PhysRevA.90.013840>.
 38. A. Pick, A. Cerjan, D. Liu, et al., “Ab Initio Multimode Linewidth Theory for Arbitrary Inhomogeneous Laser Cavities,” *Physical Review A* 91 (Jun. 2015): 063806, <https://doi.org/10.1103/PhysRevA.91.063806>.
 39. J. Andreasen, A. A. Asatryan, L. C. Botten, et al., “Modes of Random Lasers,” *Advances in Optics and Photonics* 3, no. 1 (2011): 88–127.
 40. S. Esterhazy, D. Liu, M. Liertzer, A. Cerjan, L. Ge, K. G. Makris, A. D. Stone, J. M. Melenk, S. G. Johnson, and S. Rotter, “Scalable Numerical Approach for the Steady-State Ab-Initio Laser Theory,” *Physical Review A* 90 (2014): 023816.
 41. A. Griewank and A. Walther, *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation, Second Edition* (Philadelphia, PA: SIAM, 2008).
 42. T. W. Hughes, M. Minkov, I. A. D. Williamson, and S. Fan, “Adjoint Method and Inverse Design for Nonlinear Nanophotonic Devices,” *ACS Photonics* 5, no. 12 (2018): 4781–4787.
 43. A. Cerjan, B. Redding, L. Ge, S. F. Liew, H. Cao, and A. D. Stone, “Controlling Mode Competition by Tailoring the Spatial Pump Distribution in a Laser: A Resonance-Based Approach,” *Optics Express* 24, no. 23 (Nov. 2016): 26006–26015, <https://opg.optica.org/oe/abstract.cfm?URI=oe-24-23-26006>.
 44. A. Raman and S. Fan, “Perturbation Theory for Plasmonic Modulation and Sensing,” *Physical Review B* 83, no. 20 (May 2011), <http://doi.org/10.1103/PhysRevB.83.205131>.
 45. R. E. Christiansen, P. T. Kristensen, J. Mørk, and O. Sigmund, “Impact of Figures of Merit in Photonic Inverse Design,” *Optics Express* 31, no. 5 (Feb. 2023): 8363–8374, <https://opg.optica.org/oe/abstract.cfm?URI=oe-31-5-8363>.
 46. W. Ma, R. Pestourie, and S. G. Johnson, “Time-Reversal-Symmetry Bounds on Electromagnetic Fields,” *Physical Review A* 111 (2025): 063518.

47. O. Sigmund and K. Maute, "Topology Optimization Approaches: A Comparative Review," *Structural and Multidisciplinary Optimization* 48, no. 6 (2013): 1031–1055.
48. F. Wang, B. S. Lazarov, and O. Sigmund, "On Projection Methods, Convergence and Robust Formulations in Topology Optimization," *Structural and Multidisciplinary Optimization* 43 (2011): 767–784, <https://api.semanticscholar.org/CorpusID:121060929>.
49. B. S. Lazarov and O. Sigmund, "Filters in Topology Optimization Based on Helmholtz-Type Differential Equations," *International Journal for Numerical Methods in Engineering* 86, no. 6 (2011): 765–781.
50. O. Sigmund, "Manufacturing Tolerant Topology Optimization," *Acta Mechanica Sinica* 25, no. 2 (2009): 227–239.
51. A. M. Hammond, A. Oskooi, S. G. Johnson, and S. E. Ralph, "Photonic Topology Optimization with Semiconductor-Foundry Design-Rule Constraints," *Optics Express* 29 (2021): 23916–23938.
52. M. Zhou, B. S. Lazarov, F. Wang, and O. Sigmund, "Minimum Length Scale in Topology Optimization by Geometric Constraints," *Computer Methods in Applied Mechanics and Engineering* 293 (2015): 266–282.
53. A. M. Hammond, A. Oskooi, M. Chen, Z. Lin, S. G. Johnson, and S. E. Ralph, "High-Performance Hybrid Time/Frequency-Domain Topology Optimization for Large-Scale Photonics Inverse Design," *Optics Express* 30, no. 3 (Jan. 2022): 4467–4491.
54. A. Taflov and S. Hagness, *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, 3rd ed (Artech, 2005).
55. J. Berenger, "A Perfectly Matched Layer for the Absorption of Electromagnetic Waves," *Journal of Computational Physics* 114, no. 2 (1994): 185–200.
56. J.-M. Jin, *The Finite Element Method in Electromagnetics*, 3rd ed (Wiley-IEEE Press, 2014).
57. M. Chen, R. E. Christiansen, J. A. Fan, et al., "Validation and Characterization of Algorithms and Software for Photonics Inverse Design," *Journal of the Optical Society of America B* 41, no. 2 (2024): A161–A176.
58. A. Dalklint, R. E. Christiansen, and O. Sigmund, "On Performance Bounds for Topology Optimization," (2024).
59. H. Choi, M. Heuck, and D. Englund, "Self-Similar Nanocavity Design with Ultrasmall Mode Volume for Single-Photon Nonlinearities," *Physical Review Letters* 118, no. 22 (2017): 223605.
60. M. Albrechtsen, B. V. Lahijani, and S. Stobbe, "Two Regimes of Confinement in Photonic Nanocavities: Bulk Confinement Versus Lightning Rods," *Optics Express* 30, no. 9 (2022): 15458–15469.
61. J. Andersen and V. Solodukhov, "Field Behavior Near a Dielectric Wedge," *IEEE Transactions on Antennas and Propagation* 26, no. 4 (1978): 598–602.
62. X. Qian and O. Sigmund, "Topological Design of Electromechanical Actuators with Robustness Toward Over- and Under-Etching," *Computer Methods in Applied Mechanics and Engineering* 253 (2013): 237–251.
63. M. Saritas and H. McKell, "Diffusion Length Studies in Silicon by the Surface Photovoltage Method," *Solid-State Electronics* 31, no. 5 (1988): 835–842.
64. M. Levinshtein, S. Rumyantsev, and M. Shur, *Handbook Series on Semiconductor Parameters. Volume 1: Si, Ge, C (Diamond), GaAs, GaP, GaSb, InAs, InP, InSb* (World Scientific, 1996).
65. A. David, "Long-Range Carrier Diffusion in (In,Ga)N Quantum Wells and Implications from Fundamentals to Devices," *Physical Review Applied* 15 (2021): 054015.
66. A. Fiore, M. Rossetti, B. Alloing, et al., "Carrier Diffusion in Low-Dimensional Semiconductors: A Comparison of Quantum Wells, Disordered Quantum Wells, and Quantum Dots," *Physical Review B* 70 (2004): 205311.
67. A. E. Siegman, *Lasers* (University Science Books, 1986).
68. W. W. Chow and S. Koch, *Semiconductor-Laser Fundamentals: Physics of the Gain Materials* (Springer, 1999).
69. E. Gehrig and O. Hess, "Spatio-Temporal Dynamics and Quantum Fluctuations in Semiconductor Lasers," *Springer Tracts in Modern Physics*, Springer Berlin, Heidelberg 189 (01 2003).
70. Y. Yu, A. Sakanas, A. Zali, E. Semenova, K. Yvind, and J. Mørk, "Ultra-Coherent Fano Laser Based on a Bound State in the Continuum," *Nature Photonics* 15, no. 10 (2021): 758–764.
71. P. Holewa, D. Vajner, E. Zięba-Ost'oj, et al., "High-Throughput Quantum Photonic Devices Emitting Indistinguishable Photons in the Telecom C-Band," *Nature Communications* 15 (2024): 3358.
72. P. Iyer, S. Prescott, S. Addamane, et al., "Control of Quantized Spontaneous Emission from Single GaAs Quantum Dots Embedded in Huygens' Metasurfaces," *Nano Letters* 24 (2024): 4749–4757.
73. R. E. Evans, M. K. Bhaskar, D. D. Sukachev, et al., "Photon-Mediated Interactions Between Quantum Emitters in a Diamond Nanocavity," *Science* 362, no. 6415 (2018): 662–665.
74. P.-J. Stas, Y. Q. Huan, B. Machielse, et al., "Robust Multi-Qubit Quantum Network Node with Integrated Error Detection," *Science* 378, no. 6619 (2022): 557–560.
75. A. Sipahigil, R. E. Evans, D. D. Sukachev, et al., "An Integrated Diamond Nanophotonics Platform for Quantum-Optical Networks," *Science* 354, no. 6314 (2016): 847–850.
76. "Comsol Multiphysics v. 6.2. COMSOL AB, Stockholm, Sweden," <https://www.comsol.com>.
77. K. Svanberg, "A Class of Globally Convergent Optimization Methods Based on Conservative Convex Separable Approximations," *SIAM Journal on Optimization* 12, no. 2 (2002): 555–573.
78. DTU Computing Center, "DTU Computing Center Resources," (2024), <https://doi.org/10.48714/DTU.HPC.0001>.

Appendix A: Comparison with SPA-SALT

A.1 | Single Mode Lasing Threshold

In this appendix, we derive the connection between the single pole threshold approximation in Section 2.2 and SPA-SALT. In SPA-SALT notation, Let a passive cavity resonance $\mathbf{e}_1(\mathbf{r})$ satisfy an equation similar to Equation (4), but without the gain medium,

$$\left[(\nabla \times \nabla \times) - \epsilon_c(\mathbf{r}) \left(\frac{\tilde{\omega}_1^2}{c^2} \right) \right] \mathbf{e}_1(\mathbf{r}) = 0. \quad (\text{A1})$$

Here, $\tilde{\omega}_1$ is the complex frequency of the passive cavity resonance. Under the single-pole approximation, the lasing mode $\mathbf{E}_L(\mathbf{r}) \approx a_1 \mathbf{e}_1(\mathbf{r}) / \sqrt{\int_{\Omega} \epsilon_c(\mathbf{r}) |\mathbf{e}_1(\mathbf{r})|^2 d\Omega}$, where a_1 is the complex amplitude of the field for the first lasing mode. Thus, multiplying Equation (A1) by a_1 and equating Equations (4, A1),

$$\epsilon_c(\mathbf{r}) \left(\frac{\tilde{\omega}_1^2}{c^2} \right) \mathbf{e}_1(\mathbf{r}) = (\epsilon_c(\mathbf{r}) - i\Delta\epsilon_{\mathfrak{N}}(\mathbf{r})) \left(\frac{\omega_L^2}{c^2} \right) \mathbf{e}_1(\mathbf{r}). \quad (\text{A2})$$

Rearranging,

$$\epsilon_c(\mathbf{r}) \left(\frac{\tilde{\omega}_1^2 - \omega_L^2}{\omega_L^2} \right) \mathbf{e}_1(\mathbf{r}) = -i\Delta\epsilon_{\mathfrak{N}}(\mathbf{r}) \mathbf{e}_1(\mathbf{r}), \quad (\text{A3})$$

The SPA-SALT lasing threshold is found by assuming that there is no hole-burning term in the gain medium and allowing the lasing frequency to be

different from the gain's central frequency, such that

$$\Delta\epsilon_{\mathfrak{S}}(\mathbf{r}) = \left(\frac{\gamma_{\perp}}{\gamma_{\perp} - i(\omega_L - \omega_a)} \right) \frac{D_0(\mathbf{r})d}{1 + \Gamma_1 e_c^{-2} |a_1|^2 |\mathbf{e}_1(\mathbf{r})|^2 / \int_{\Omega} \epsilon_c(\mathbf{r}) |\mathbf{e}_1(\mathbf{r})|^2 d\Omega}, \quad (\text{A4})$$

becomes

$$\Delta\epsilon_{\mathfrak{S}}^{(\text{threshold})}(\mathbf{r}) = \left(\frac{\gamma_{\perp}}{\gamma_{\perp} - i(\omega_L - \omega_a)} \right) D_0(\mathbf{r})d, \quad (\text{A5})$$

where as a reminder $\Gamma_1 = \gamma_{\perp}^2 / [\gamma_{\perp}^2 + (\omega_L - \omega_a)^2]$. Altogether, at the first lasing threshold, this yields

$$\epsilon_c(\mathbf{r}) \left(\frac{\tilde{\omega}_1^2 - \omega_L^2}{\omega_L^2} \right) \mathbf{e}_1(\mathbf{r}) = -i \left(\frac{\gamma_{\perp}}{\gamma_{\perp} - i(\omega_L - \omega_a)} \right) d_{\text{threshold}} D_0(\mathbf{r}) \mathbf{e}_1(\mathbf{r}). \quad (\text{A6})$$

The SPA-SALT threshold prediction [2, 43] is then found by integrating both sides with respect to $\mathbf{e}_1(\mathbf{r})$ as SALT generally uses the biorthogonal normalization of the system's lasing modes and resonances. We can instead recover Equation (9) by restoring the assumption that $\omega_L = \omega_a$, $\tilde{\omega}_1 = \omega_L - i\gamma_1$, and expanding $(\tilde{\omega}_1^2 - \omega_L^2) / \omega_L^2$ to first order in γ_1 , yielding

$$\epsilon_c(\mathbf{r}) \left(\frac{-2i\gamma_1}{\omega_L} \right) \mathbf{e}_1(\mathbf{r}) = -id_{\text{threshold}} D_0(\mathbf{r}) \mathbf{e}_1(\mathbf{r}), \quad (\text{A7})$$

which is then integrated by $\mathbf{e}_1^*(\mathbf{r})$,

$$d_{\text{threshold}} = \left(\frac{2\gamma_1}{\omega_L} \right) \left(\frac{\int_{\Omega} \mathbf{e}_1^*(\mathbf{r}) \epsilon_c(\mathbf{r}) \mathbf{e}_1(\mathbf{r}) d\Omega}{\int_{\Omega} \mathbf{e}_1^*(\mathbf{r}) D_0(\mathbf{r}) \mathbf{e}_1(\mathbf{r}) d\Omega} \right). \quad (\text{A8})$$

A.2 | Interacting Threshold for a Second Lasing Mode

To estimate the effect of spatial hole-burning on the thresholds of other potential lasing modes, we again approximate the spatial profile of each potential mode to be that of a different cavity resonance, $\mathbf{e}_m(\mathbf{r})$. We note that while the approximation of treating the first lasing mode as having the same spatial profile of a high- Q resonance is quite good when the passive cavity resonance frequency is well-aligned with the central frequency of the gain (as we do here, see Appendix D), this same approximation on other lasing modes may be worse, especially if the resonance frequencies are outside of the gain window and thus experience significant line-pulling. Nevertheless, for the systems we consider here, which are tailored to support a single high- Q passive cavity resonance, this approximation is still sufficient to estimate the range of single-mode behavior.

For any second lasing mode to reach threshold, such that the total lasing field can be decomposed into two different modes with distinct lasing frequencies $\mathbf{E}_L(\mathbf{r}; t) \approx a_1 \mathbf{e}_1(\mathbf{r}) e^{-i\omega_{L1} t} + a_2 \mathbf{e}_2(\mathbf{r}) e^{-i\omega_{L2} t}$, we seek a nontrivial (i.e., for $|a_2| \ll 1$, but $|a_2| \neq 0$) solution to

$$\begin{aligned} \epsilon_c(\mathbf{r}) \left(\frac{\tilde{\omega}_2^2 - \omega_{L2}^2}{\omega_{L2}^2} \right) \mathbf{e}_2(\mathbf{r}) = & -i \left(\frac{\gamma_{\perp}}{\gamma_{\perp} - i(\omega_{L2} - \omega_a)} \right) \\ & \times \left(\frac{D_0(\mathbf{r}) d_{\text{int thr}}^{(2)}}{1 + \Gamma_1 e_c^{-2} |a_1|^2 |\mathbf{e}_1(\mathbf{r})|^2 / \int_{\Omega} \epsilon_c(\mathbf{r}) |\mathbf{e}_1(\mathbf{r})|^2 d\Omega} \right) \\ & \times \mathbf{e}_2(\mathbf{r}). \end{aligned} \quad (\text{A9})$$

Note that if $|a_1| = 0$, the non-interacting threshold for this other mode is given by

$$d_{\text{thr}}^{(2)} = i \left(\frac{\tilde{\omega}_2^2 - \omega_{L2}^2}{\omega_{L2}^2} \right) \left(\frac{\gamma_{\perp} - i(\omega_{L2} - \omega_a)}{\gamma_{\perp}} \right) \left(\frac{\int_{\Omega} \mathbf{e}_2^*(\mathbf{r}) \epsilon_c(\mathbf{r}) \mathbf{e}_2(\mathbf{r}) d\Omega}{\int_{\Omega} \mathbf{e}_2^*(\mathbf{r}) D_0(\mathbf{r}) \mathbf{e}_2(\mathbf{r}) d\Omega} \right), \quad (\text{A10})$$

thus

$$d_{\text{int thr}}^{(2)} = d_{\text{thr}}^{(2)} \left[1 + \Gamma_1 e_c^{-2} |a_1|^2 \frac{\int_{\Omega} \epsilon_c(\mathbf{r}) |\mathbf{e}_2(\mathbf{r})|^2 |\mathbf{e}_1(\mathbf{r})|^2 d\Omega}{\left(\int_{\Omega} \epsilon_c(\mathbf{r}) |\mathbf{e}_2(\mathbf{r})|^2 d\Omega \right) \left(\int_{\Omega} \epsilon_c(\mathbf{r}) |\mathbf{e}_1(\mathbf{r})|^2 d\Omega \right)} \right]. \quad (\text{A11})$$

To completely specify the interacting threshold of a second lasing mode within the SPA-SALT approximation (though noting that here we are using the Hermitian normalization and approximate orthogonality), we can solve for $|a_1|$ at arbitrary d in the single mode regime up to and including the second interacting threshold using Equations A4 and A5 as

$$\frac{d}{d_{\text{threshold}}} - 1 = \Gamma_1 e_c^{-2} |a_1|^2 \left[\frac{\int_{\Omega} \epsilon_c(\mathbf{r}) |\mathbf{e}_1(\mathbf{r})|^2 |\mathbf{e}_2(\mathbf{r})|^2 d\Omega}{\left(\int_{\Omega} \epsilon_c(\mathbf{r}) |\mathbf{e}_1(\mathbf{r})|^2 d\Omega \right)^2} \right]. \quad (\text{A12})$$

To proceed, let

$$\chi_{mn} = \frac{\int_{\Omega} \epsilon_c(\mathbf{r}) |\mathbf{e}_m(\mathbf{r})|^2 |\mathbf{e}_n(\mathbf{r})|^2 d\Omega}{\left(\int_{\Omega} \epsilon_c(\mathbf{r}) |\mathbf{e}_m(\mathbf{r})|^2 d\Omega \right) \left(\int_{\Omega} \epsilon_c(\mathbf{r}) |\mathbf{e}_n(\mathbf{r})|^2 d\Omega \right)} \quad (\text{A13})$$

denote the interaction coefficients [2]. Solving Equation (A12) for $\Gamma_1 e_c^{-2} |a_1|^2$ and into Equation (A11), one can write an implicit equation for the interacting second modal threshold,

$$d_{\text{int thr}}^{(2)} = d_{\text{thr}}^{(2)} \left(1 + \left(\frac{d_{\text{int thr}}^{(2)}}{d_{\text{threshold}}} - 1 \right) \frac{\chi_{21}}{\chi_{11}} \right), \quad (\text{A14})$$

where χ_{11} denotes the self-saturation of the first mode (how much it depletes its own gain) and χ_{12} denotes the cross-saturation of mode 2 due to mode 1 (caused by the mode overlap). Finally, one can recover an explicit equation for the interacting thresholds of a second lasing mode as

$$d_{\text{int thr}}^{(2)} = d_{\text{thr}}^{(2)} \left(\frac{1 - \chi_{21} / \chi_{11}}{1 - d_{\text{thr}}^{(2)} \chi_{21} / (d_{\text{threshold}} \chi_{11})} \right). \quad (\text{A15})$$

Note that this formula for the interacting threshold must be used with some caution, as without some care it can produce the misleading result of a negative interacting threshold. Instead, such a negative $d_{\text{int thr}}^{(2)}$ indicates that the mode never reaches threshold. To see, this, first note that by definition $d_{\text{thr}}^{(2)} > d_{\text{threshold}}$, and similarly, $\chi_{21} \leq \chi_{11}$, as the interacting modal overlap between two modes cannot be greater than the interacting overlap of a mode with itself. If $\chi_{21} = 0$, i.e., the two modes do not spatially overlap, the interacting threshold of the second lasing mode is the same as its non-interacting value, $d_{\text{int thr}}^{(2)} = d_{\text{thr}}^{(2)}$. As χ_{21} is increased from zero, the ratio of $d_{\text{thr}}^{(2)} / d_{\text{threshold}}$ that yields a positive denominator becomes bounded. If, as a thought experiment, one could independently control $d_{\text{thr}}^{(2)}$ without altering χ_{21} , increasing $d_{\text{thr}}^{(2)}$ above $d_{\text{threshold}}$ first yields a zero denominator, indicating a diverging $d_{\text{int thr}}^{(2)}$, before the interacting threshold becomes negative. Thus, within the SPA-SALT approximation,

$$\frac{d_{\text{thr}}^{(2)} \chi_{21}}{d_{\text{threshold}} \chi_{11}} \geq 1 \Rightarrow d_{\text{int thr}}^{(2)} \rightarrow \infty. \quad (\text{A16})$$

This divergence arises from the SPA-SALT approximation, which fixes the modal frequencies and spatial profiles. Under this assumption, a second mode cannot reshape its profile to access unused gain, so the interacting threshold formally diverges. Physically, this does not mean the threshold is infinite; it simply indicates that the second mode would

require pumping far beyond the nonlinear regime where SPA-SALT is valid. In reality, the threshold is extremely large but finite, supporting the conclusion that the system remains effectively single-mode under typical operating conditions. In Appendix D we use the interacting thresholds to see compare the lasing thresholds of the first and second lasing modes.

A.3 | Scalar Residual for the SALT Model

Using the expression in Equation (A12) one can derive the lasing intensity for an arbitrary pumping strength above threshold $\Delta d = d - d_{\text{thresh}}$,

$$|a_1|^2 = \frac{\Delta d}{\Gamma_1 e_c^{-2} \chi_{11} d_{\text{thresh}}}, \quad (\text{A17})$$

which is the amplitude expression that we consider in Appendix B close to threshold ($\Delta d \ll 1$) and in the weak pumping regime ($|a_1|^2 \ll 1$).

By using the expression for the lasing intensity derived in Equation (A17) together with Equation (3) in the single mode regime, it is possible to find the population inversion above threshold,

$$D(\mathbf{r}) = \left[1 + \frac{d - d_{\text{thresh}}}{\chi_{11} d_{\text{thresh}}} |\mathbf{e}_1(\mathbf{r})|^2 \right]^{-1} D_0(\mathbf{r}) d, \quad (\text{A18})$$

which can be directly inserted into Equation (2). With this, we can define a normalized scalar residual measure based on the weak form of our finite element formulation that is a function of the pumping strength as

$$r(d) = \frac{\int_{\Omega} |\nabla \times \mathbf{E}_r(\mathbf{r})|^2 - [\varepsilon_c(\mathbf{r}) - iD(\mathbf{r}, d)](\omega_1/c)^2 |\mathbf{E}_r(\mathbf{r})|^2 d\Omega}{\int_{\Omega} (\omega_1/c)^2 \varepsilon_c(\mathbf{r}) |\mathbf{E}_r(\mathbf{r})|^2 d\Omega}, \quad (\text{A19})$$

where we approximate the passive cavity mode with the reciprocal field in the high- Q limit, while still considering the lasing frequency to be the real part of the cavity's resonance frequency ($\omega_L \approx \omega_1$). Thus, the residual will account for errors in differences between the frequency of the mode and the one used to solve the reciprocal problem, as well as the error introduced through the nonlinear hole burning term. In Appendix D we evaluate this residual for different values of the pumping strength d to validate the high- Q approximation at threshold and above, which allows to see how spatial hole burning affects the designs optimized close to threshold ($\Delta d = d - d_{\text{thresh}} \ll 1$).

Appendix B: Lasing Mode Amplitude in the Weak Pumping Regime

Following a similar derivation to the single mode lasing amplitude in Appendix A, in this appendix we detail the calculation of the lasing amplitude in the high- Q , small amplitude, and small pumping strength limit. It is possible to expand Equation (7) for small amplitudes ($|a_1|^2 \ll 1$), including the zeroth and the first-order terms. This gives

$$\Delta \varepsilon_{\mathfrak{G}}(\mathbf{r}) \approx D_0(\mathbf{r}) d \left(1 - |a_1|^2 \frac{e_c^{-2} |\mathbf{e}_1(\mathbf{r})|^2}{\int_{\Omega} \varepsilon_c(\mathbf{r}) |\mathbf{e}_1(\mathbf{r})|^2 d\Omega} \right). \quad (\text{B1})$$

Note that this first-order correction has the form of a Kerr nonlinearity, where the change in dielectric permittivity depends on the electric-field intensity. In the high- Q cavity limit, we can use Equation (6), where we use perturbation theory to calculate the shift in frequency of the mode, and substitute in the lasing condition in Equation (8). Using these expressions and assuming that we operate close to the lasing threshold where $d \approx d_{\text{thresh}}$ and $\Delta d = d - d_{\text{thresh}} \ll d_{\text{thresh}}$, the lasing mode intensity is given by

$$|a_1|^2 \approx \frac{\Delta d}{d_{\text{thresh}}} \frac{\int_{\Omega} \varepsilon_c(\mathbf{r}) |\mathbf{e}_1(\mathbf{r})|^2 d\Omega \int_{\Omega} D_0(\mathbf{r}) |\mathbf{e}_1(\mathbf{r})|^2 d\Omega}{e_c^{-2} \int_{\Omega} D_0(\mathbf{r}) |\mathbf{e}_1(\mathbf{r})|^4 d\Omega}, \quad (\text{B2})$$

where Δd denotes difference of the pumping strength d with the threshold. (At large amplitudes, many additional nonlinear effects arise, including competition from additional lasing modes, eventually requiring a full nonlinear solver [40].)

Appendix C: Lasing Threshold and FOM With Steady-state Diffusion

In this appendix, we revisit the derivations of the laser properties (e.g., threshold, lasing mode intensity) and FOM, including steady-state diffusion effects through the C-SALT formalism [4]. We introduce diffusion in the gain model in Equation (16). By rewriting the lasing mode as a function of the passive cavity mode, we find

$$\Delta \varepsilon_{\mathfrak{G}}(\mathbf{r}) = \mathbb{S} \left(\mathbb{I} + |a_1|^2 \frac{e_c^{-2} |\mathbf{e}_1(\mathbf{r})|^2}{\int_{\Omega} \varepsilon(\mathbf{r}) |\mathbf{e}_1(\mathbf{r})|^2 d\Omega} \mathbb{S} \right)^{-1} [D_0](\mathbf{r}) d. \quad (\text{C1})$$

Similar to the case without diffusion, close to the lasing threshold, we assume a small amplitude ($|a_1|^2 \ll 1$), so we can expand the expression in a Taylor series around $|a_1|^2 = 0$. For zero amplitude the change in permittivity is $\Delta \varepsilon_{\mathfrak{G}}(\mathbf{r}) = \mathbb{S}[D_0](\mathbf{r}) d$. Using this expression together with the lasing condition in Equation (8), gives an updated version of the threshold pumping strength that accounts for diffusion effects, resulting in Equation (17).

By including the zero-th and the first-order terms in the expansion, we calculate the change in permittivity in the gain medium

$$\Delta \varepsilon_{\mathfrak{G}}(\mathbf{r}) \approx \mathbb{S} \left(\mathbb{I} - |a_1|^2 \frac{e_c^{-2} |\mathbf{e}_1(\mathbf{r})|^2}{\int_{\Omega} \varepsilon(\mathbf{r}) |\mathbf{e}_1(\mathbf{r})|^2 d\Omega} \mathbb{S} \right) [D_0](\mathbf{r}) d. \quad (\text{C2})$$

Following the same relations as in Appendix B gives an expression for the intensity of the lasing mode

$$|a_1|^2 \approx \frac{\Delta d}{d_{\text{thresh}}} \frac{\int_{\Omega} \varepsilon(\mathbf{r}) |\mathbf{e}_1(\mathbf{r})|^2 d\Omega \int_{\Omega} \mathbb{S}[D_0](\mathbf{r}) |\mathbf{e}_1(\mathbf{r})|^2 d\Omega}{e_c^{-2} \int_{\Omega} \mathbb{S}[|\mathbf{e}_1(\mathbf{r})|^2 \mathbb{S}[D_0]](\mathbf{r}) |\mathbf{e}_1(\mathbf{r})|^2 d\Omega}. \quad (\text{C3})$$

Using this expression together with the TCMT equations (Section 2.3), we obtain the FOM of Equation (18).

Appendix D: Validating the SPA-SALT Approximations

To validate the SPA-SALT approximation, we will use the derivations in Appendix A. Primarily, we will evaluate the two key approximations in our SPA-SALT-based optimization framework: that a single cavity mode will lase with a threshold significantly lower than other cavity modes, and that the electric field in the reciprocal problem accurately represents the single lasing mode at threshold and above. As an example, we will evaluate this by using the device optimized in 2d for a gain region with $\sigma = 500$ nm (i.e., the cavity design shown in Figure 3d and Figure 4).

D.1 | Lasing Threshold for a Second Lasing Mode

First, we identify the cavity modes closest to the optimized lasing mode. These modes are resonances with a maximum Q -factor of $Q \approx 200$ (≈ 1 -2 orders of magnitude lower than the lasing mode), which are spectrally separated by just a few nanometers. Given this spectral closeness, we approximate that the closest modes do not experience any line pulling effects ($\gamma_1 \rightarrow \infty$, so that $\Gamma_1 \rightarrow 1$ and $\Gamma_2 \rightarrow 1$). By using the expressions for the interacting threshold of a second lasing mode in Equation (A15) and assuming $\omega_{L_2} = \omega_2$, we find that the value for the second interacting threshold for the closest modes diverge: $d_{\text{int thr}}^{(2)} \rightarrow \infty$. This divergence

TABLE E1 | Continuation scheme parameters in the topology optimization framework.

Continuation step	0	1	2	3	4	5	6	7	8
Iteration	0	100	200	300	400	500	600	700	800
Threshold sharpness (β)	2.5	5	7.5	10	15	25	50	75	100
Attenuation factor (α')	0.0	0.0	0.01	0.1	0.2	0.4	0.8	0.8	0.8
Artificial loss (α)	0.1	0.05	0.025	0.01	$5 \cdot 10^{-3}$	0.0	0.0	0.0	0.0
Length scale error (ϵ)	1	1	1	10^{-3}	$5 \cdot 10^{-4}$	10^{-4}	$5 \cdot 10^{-5}$	10^{-5}	$5 \cdot 10^{-5}$

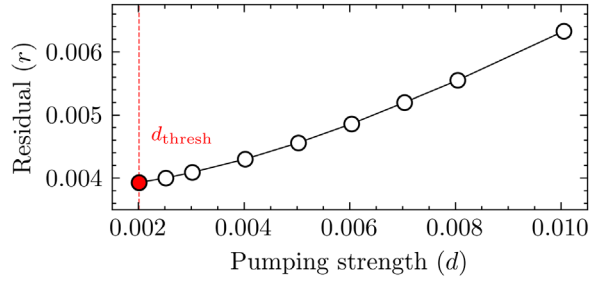


FIGURE D1 | Validation of the linear response high-Q approximation at and above threshold. The residual as a function of pump strength d for pumping strengths up to $d \approx 5d_{\text{thresh}}$. The lasing threshold is marked in red at $d_{\text{thresh}} = 2.01 \cdot 10^{-3}$.

arises from the SPA-SALT approximation, which assumes that the modal frequencies and spatial profiles are fixed. Under this assumption, a second mode cannot adapt its profile to access unused gain, so the interacting threshold formally diverges. Physically, this does not imply an infinite threshold; rather, the second mode would require pumping far beyond the nonlinear regime where SPA-SALT is valid, so the threshold is extremely large but finite. This means that within the SPA-SALT approximation a second mode will not reach threshold in the presence of modal interactions. This result thus supports the assumption that our system is dominated by a single lasing mode. Furthermore, the small modal overlap (quantified by χ_{21}/χ_{11}) ensures that even modes close in frequency are strongly suppressed by the first lasing mode, reinforcing single-mode operation. In a full SALT calculation allowing modal profile adaptation, the second mode could eventually turn on at very high pump, but this occurs well outside the parameter regime considered here.

D.2 | Linear Response Validation for High-Q Cavities

Next, we validate the error between the linear response of the reciprocal system and the SALT model above the threshold by using the scalar residual measure defined in Equation (A19). The residual gives a scalar measure of the relative error between the linear solution and the SALT model, which can be used to validate the high-Q approximation (i.e., that the reciprocal field is proportional to the cavity mode).

To show the validity of this approximation in Figure D1, we plot the residual (Equation (A19)) as a function of the scalar pumping strength (d). Close to the threshold (d_{thresh}), the residual is very low (a relative error of $\lesssim 0.4\%$), whereas as the pumping strength increases, the residual and the relative error start to increase. Nevertheless, even for a larger pumping strength of $d \approx 5d_{\text{thresh}}$, the relative error is still smaller than $\lesssim 1\%$. We note that for even larger pumping strengths, the residual grows as a linear function, and we calculate that for a pumping strength of $d \approx 100d_{\text{thresh}}$ the relative error is still $\approx 5\%$. These results confirm the robustness and validity of the approximation for the inverse-designed high-Q nanolaser cavities.

Appendix E: Simulation and Optimization Parameters

In this section we detail the parameters used in the simulation and optimization of the nanolaser devices. The numerical model is discretized using the finite-element method [56], with first-order Nedgelec elements [56] for the frequency-domain electromagnetic problem and linear Lagrange elements for the steady-state exponentially damped diffusion problems.

The simulations are performed in COMSOL Multiphysics (version 6.2) [76] and the optimization is solved using the Globally Convergent Method of Moving Asymptotes (GCMMA) [77]. All computations are executed on the DTU Computing Center HPC cluster [78].

Geometrical Parameters

Two dimensional systems: In the 2d computational results (Section 3.1, 3.2) we consider a geometry based on a rectangular simulation domain with length $l = 13.95 \mu\text{m}$ (along the x direction) and width $w = 4.55 \mu\text{m}$ (along the y direction), where the rectangular design domain (Ω) is centered at position $\mathbf{r}_0 = (1.1625 \mu\text{m}, 0 \mu\text{m})$, and has side of length of $l_\Omega = 1.5l = 2.325 \mu\text{m}$. The design domain is connected to a waveguide with width $w_{\text{wg}} = 500 \text{ nm}$, which is excited at position $\mathbf{r}_{\text{src}} = (-1 \mu\text{m}, 0 \mu\text{m})$.

Three dimensional systems: In the 3d computational results (Section 3.3), we extend the previous geometry by extruding the 2D design domain (Ω) and waveguide out-of-plane (along the z -direction) to a height of $h_\Omega = 220 \text{ nm}$, and placing the resulting structure on a silica substrate. The design domain is at the center of a simulation domain with total height of $h = 1.5l = 2.325 \mu\text{m}$.

Numerical Discretization

In the 2d plane the design is meshed with a structured square grid with 40 nm side length, while the rest of the geometry is discretized using second-order elements on an unstructured triangular mesh with a side length of 40 nm in the solid (n_1) and 155 nm in the air (n_0), so that the size corresponds to roughly $1/10$ in-material wavelengths. In the 3d simulations the waveguide and design domain are extruded with 5 elements out-of-plane, and the rest of the simulation domain is meshed with tetrahedral elements with side lengths corresponding to $1/10$ in-material wavelengths. Note that all reported performances and fields are evaluated using the final post processed design, discretized using an unstructured triangular (tetrahedral) body fitted mesh with 20 nm ($\sim \lambda/(20 n_1)$) side length for the optimized structures.

Optimization Parameters

The main optimization parameter choices are the threshold value $\eta = 0.5$ and filter radius $r_f = 100 \text{ nm}$. The geometric length scale constraints [52]

are enforced using the parameters $c = 39$, and $\eta_e = 0.75$ and $\eta_d = 0.25$, as to ensure a lengthscale $\sim r_f/2 = 50$ nm [53]. The rest of the parameters, including the length scale error (ϵ) needed to enforce the length scale constraints [52], are chosen as part of a continuation scheme, which is essential to enable the use of gradient-based methods and ensures a well-performing final binary design. In the continuation scheme we vary the simulation parameters as a function of the iteration number to find a binary and well performing final design. We summarize those parameter choices in Table E1.