Supplementary Information

Inverse design of open nanophotonic cavities
through differentiable mode expansions

Author list and affiliations in preparation

September 5, 2026

This document is self-contained. Sections S1–S13 develop the theory and the numerical construction in one consistent notation, starting from Maxwell's equations and ending at the optimizer. Sections S14–S19 present the evidence: analytical convergence and gradient validation, the independent-FDTD protocol and its ringdown audit, the two device campaigns in full, a catalogue of rejected approximations and outright failures, and a reproducibility and provenance appendix that maps every theory section onto the retained derivation record. Every numerical value printed here is generated from a machine-readable record; none is transcribed by hand.

Contents

S1 Scope, notation, and the claim ledger . . . 4
- S1.1 What this document contains . . . 4
- S1.2 Two observables, never interchanged . . . 4
- S1.3 Claim ledger . . . 4
- S1.4 Notation . . . 5

S2 Conventions, quasinormal modes, and normalization . . . 5
- S2.1 Time convention and the pole . . . 5
- S2.2 Nonlinear eigenproblem and biorthogonality . . . 5
- S2.3 Normalization . . . 6
- S2.4 The logarithmic objective . . . 6

S3 Exact parameterized dielectric and its transforms . . . 6
- S3.1 Reference and contrast . . . 6
- S3.2 Closed-form Fourier factors . . . 7
- S3.3 Analytic geometry derivatives . . . 7
- S3.4 What this parameterization buys . . . 8

S4 Outgoing integral equation and the regularized determinant . . . 8
- S4.1 Reduction to the holes . . . 8
- S4.2 Why the determinant must be regularized . . . 8
- S4.3 Finite sequence . . . 9
- S4.4 Relation to cell scattering matrices . . . 9

S5 Reference Green tensor: homogeneous cladding . . . 9
- S5.1 Fourier reduction along the beam . . . 9
- S5.2 Exact homogeneous symbol . . . 10
- S5.3 Cross-section basis and Galerkin matrices . . . 10
- S5.4 Dyson resummation of the core . . . 10
- S5.5 Guided poles are not inserted by hand . . . 11
- S5.6 On-shell evaluation inside the light cone . . . 11

S6 Reference Green tensor: stratified stack . . . 11

S6.1 Why the homogeneous reference fails here . . . 11
S6.2 Generalized Fresnel recursion . . . 12
S6.3 Scalar vertical kernels and the Wronskian . . . 12
S6.4 Adding the strip . . . 12
S7 The outgoing contour and its quadrature . . . 13
S7.1 Branch topology in scaled momentum . . . 13
S7.2 The deformation . . . 13
S7.3 Orientation test . . . 13
S7.4 Panelized quadrature . . . 13
S7.5 Known failure modes . . . 14
S8 Aperture basis and projection . . . 14
S8.1 Normalized hole coordinates . . . 14
S8.2 Zernike–Legendre vector basis . . . 14
S8.3 Exact thickness projection . . . 15
S8.4 Projection into the cross-section basis . . . 15
S8.5 Transpose, not conjugate . . . 15
S8.6 Production orders . . . 15
S9 Symmetry reduction and finite-array assembly . . . 16
S9.1 The assembled operator . . . 16
S9.2 Symmetry sectors . . . 16
S9.3 Finite array, not a periodic cell . . . 16
S9.4 Block structure and cell cascades . . . 16
S10 Feed-pole extraction and residues . . . 16
S10.1 The problem . . . 16
S10.2 Laurent extraction . . . 17
S10.3 Exact real-space restoration . . . 17
S10.4 Magnitude of the effect . . . 17
S10.5 What must be differentiated . . . 18
S11 Continuation, pole search, and mode identity . . . 18
S11.1 Locating the first pole . . . 18
S11.2 Continuation off the real axis . . . 18
S11.3 Mode identity under geometry change . . . 18
S11.4 Multiple poles and near-degeneracies . . . 19
S12 Pole and quality-factor derivatives . . . 19
S12.1 The implicit derivative . . . 19
S12.2 Assembling $\partial_{\theta_j}\mathbf{A}$ . . . 19
S12.3 Reverse-mode contraction . . . 20
S12.4 Boundary form . . . 20
S12.5 Boundary metric . . . 20
S12.6 Trust ladder . . . 21
S13 Optimizer, trust region, and acceptance gates . . . 21
S13.1 Objective and variables . . . 21
S13.2 Step control . . . 21
S13.3 Acceptance gates . . . 22
S13.4 One loop, written out . . . 22
S13.5 What the independent solver is not allowed to do . . . 22
S14 Analytical convergence and gradient validation . . . 23
S14.1 Operator identities . . . 23
S14.2 Refinement convergence . . . 23
S14.3 Gradient validation . . . 23

S14.4 A directional gate . . . 24
S14.5 Complete audited value table . . . 24
S15 Independent-FDTD protocol and the ringdown audit . . . 25
S15.1 Role and separation . . . 25
S15.2 Endpoint procedure . . . 25
S15.3 Ringdown adequacy . . . 26
S15.4 Predeclared promotion panel . . . 26
S15.5 A required amendment to the panel . . . 27
S15.6 Complete adequacy table . . . 27
S16 Air-clad Quan–Lončar campaign . . . 27
S16.1 Device and parameterization . . . 27
S16.2 A lineage caution . . . 28
S16.3 Optimization . . . 28
S16.4 Geometry outcome . . . 28
S16.5 Independent evaluation . . . 28
S16.6 Status . . . 29
S17 Multilayer Design23 campaign . . . 29
S17.1 Device and parameterization . . . 29
S17.2 Trajectory . . . 29
S17.3 The excluded observation . . . 30
S17.4 Protocol change . . . 30
S17.5 Geometry outcome . . . 31
S17.6 Spectral and modal signature . . . 31
S17.7 Independent evaluation and status . . . 31
S18 Rejected approximations and failures . . . 32
S18.1 Few-mode and closed bases . . . 32
S18.2 Unregularized collocation determinant . . . 32
S18.3 Real-axis quadrature at complex frequency . . . 32
S18.4 Hermitian conjugate in place of transpose . . . 32
S18.5 Layered cavity without feed-pole subtraction . . . 32
S18.6 The flat direction that looked like a 3.6× gain . . . 33
S18.7 Mesh-motion blockade . . . 33
S18.8 Short-ringdown quality factors . . . 33
S18.9 Objectives that have not passed their own gates . . . 33
S19 Reproducibility and provenance . . . 34
S19.1 How the numbers reach this document . . . 34
S19.2 Statistical rules declared in advance . . . 34
S19.3 Provenance of the theory sections . . . 34
S19.4 Raw artifacts . . . 35
S19.5 Build . . . 35

Theory and construction

S1 Scope, notation, and the claim ledger

S1.1 What this document contains

The main text states a construction and two device results in the space available to it. This document gives the construction in full, in one notation, and gives the evidence in full, including the parts that failed.

Sections S2–S13 are a linear development: conventions and the quasinormal-mode (QNM) normalization (§S2); the exact parameterized dielectric and its closed-form transforms (§S3); the outgoing volume-integral equation and its regularized determinant (§S4); the reference Green tensor for a homogeneous cladding (§S5) and for a stratified stack (§S6); the outgoing contour and its quadrature (§S7); the aperture basis (§S8); symmetry reduction and array assembly (§S9); feed-pole extraction (§S10); complex-frequency continuation and mode identity (§S11); the pole and Q derivatives (§S12); and the optimizer (§S13).

Sections S14–S19 are evidence: analytical convergence and gradient validation (§S14); the independent-FDTD protocol and the ringdown audit (§S15); the two device campaigns (§S16, §S17); a catalogue of rejected approximations and failures (§S18); and reproducibility with a provenance map (§S19).

S1.2 Two observables, never interchanged

Throughout, Q_{ana} denotes a quality factor computed from a pole of the analytical operator of this work, and Q_{FDTD} denotes a quality factor fitted to the ringdown of an independent finite-difference time-domain simulation. They are different observables computed by different means from different approximations of the same device. No expression in this work converts one into the other, and no plot places them on a shared axis without saying so. Where a ratio $Q_{\text{ana}}/Q_{\text{FDTD}}$ appears it is a measure of disagreement between two models, not an error in either.

S1.3 Claim ledger

Every quantitative statement in the main text carries one of three statuses, recorded in a machine-readable ledger alongside the source.

validated

The construction of the open-pole operator (§S4–§S9) and its differentiability (§S12). The supporting evidence is §S14: operator residual, complex symmetry, refinement convergence, and analytic-versus-finite-difference gradient agreement.

provisional

The two device gains, 3.8407× and 3.0666×, and the claim that both devices exhibit transfer. These are historical endpoint measurements under stated FDTD settings. They become validated only when the predeclared panel of §S15 passes with the run-time amendment recorded there.

retired

Results superseded during development and retained only as counterexamples, catalogued in §S18. The most consequential is a pre-repair analytical operator that reported a 3.6045899× improvement along a direction that independent FDTD showed to be flat.

S1.4 Notation

symbol	meaning
$\widetilde{\omega} = \omega_r + \mathrm{i}\omega_i$	complex pole, $\omega_i < 0$
$k_0 = \omega/c$, $k_c = n_c\omega/c$	vacuum and cladding wavenumbers
β	longitudinal (beam-axis) momentum
$u = \beta/k_c$	scaled longitudinal momentum
ϵ_b, ϵ_c	core and cladding permittivity
$\chi_b = \epsilon_b - \epsilon_c$	core susceptibility contrast
W, T	beam width and patterned-layer thickness
$x_j, r_{x,j}, r_{y,j}$	centre and semi-axes of hole j
$\boldsymbol{\theta}$	vector of free geometry coordinates
$\mathbf{G}_{\mathrm{c}}, \mathbf{G}_{\mathrm{wg}}, \mathbf{G}_{\mathrm{b}}$	cladding, waveguide, stratified Green tensors
$\mathbf{G}_H^+$	outgoing Green matrix projected on the holes
$\mathbf{A}(\omega, \boldsymbol{\theta})$	assembled pole operator
$\boldsymbol{u}, \boldsymbol{v}$	right and left null vectors of $\mathbf{A}$
N_{xy}, N_z, N_β	aperture, vertical, and contour orders
$\mathcal{N}$	FDTD ringdown windows per amplitude lifetime

Vectors are bold, matrices upright bold, and the transpose $(\cdot)^T$ is used wherever analytic continuation is required; the Hermitian conjugate appears only where a quantity is evaluated on the real-frequency boundary and is explicitly flagged there.

S2 Conventions, quasinormal modes, and normalization

S2.1 Time convention and the pole

All fields carry $\exp(-\mathrm{i}\omega t)$. Maxwell's equations for nonmagnetic, nondispersive, isotropic media reduce to

$$\nabla \times \nabla \times \mathbf{E} - \frac{\omega^2}{c^2}\epsilon(\mathbf{r})\mathbf{E} = 0. \tag{S1}$$

A resonance of an open system is not an eigenvalue of (S1) under square-integrable boundary conditions; it is a pole of the associated scattering problem, obtained by imposing purely outgoing behaviour and continuing ω off the real axis. With the above time convention a passive outgoing pole satisfies

$$\widetilde{\omega} = \omega_r + \mathrm{i}\omega_i, \qquad \omega_i < 0, \qquad Q = -\frac{\omega_r}{2\omega_i}. \tag{S2}$$

The sign of ω_i is not a convention to be chosen at the end of a calculation. It is a test: a construction that returns $\omega_i > 0$ for a passive dielectric has either continued onto the wrong Riemann sheet or has an incoming rather than outgoing Green function, and both errors occurred during development (§S18).

The associated field grows exponentially at large distance—the well-known QNM divergence—which is why every normalization used here is defined through the operator rather than through a field-energy integral over an unbounded region [1, 2, 3].

S2.2 Nonlinear eigenproblem and biorthogonality

Every discretization used in this work produces a finite matrix that depends nonlinearly on frequency,

$$\mathbf{A}(\widetilde{\omega}, \boldsymbol{\theta})\, \boldsymbol{u} = 0, \qquad \boldsymbol{v}^T \mathbf{A}(\widetilde{\omega}, \boldsymbol{\theta}) = 0. \tag{S3}$$

The nonlinearity is essential and not an artifact: the reference Green tensor, the radiation branch cuts, and the contour of §S7 all depend on ω. Equation (S3) is a nonlinear eigenvalue problem [4, 5], and $\boldsymbol{v}, \boldsymbol{u}$ are its left and right null vectors at a simple root of $\det \mathbf{A}$.

The pairing between $\boldsymbol{v}$ and $\boldsymbol{u}$ is the transpose bilinear form, not the Hermitian sesquilinear form. For real-valued basis functions and real ω the two coincide, which is why the error is easy to make and hard to notice; but only the bilinear form is analytic in ω, so only it survives continuation into the lower half plane. Reciprocity of the underlying Maxwell problem implies that $\mathbf{A}$ is complex symmetric, $\mathbf{A} = \mathbf{A}^T$, whence $\boldsymbol{v} = \boldsymbol{u}$ up to normalization. We do *not* impose this. We compute $\boldsymbol{v}$ and $\boldsymbol{u}$ independently and use $\|\mathbf{A} - \mathbf{A}^T\|/\|\mathbf{A}\|$ as a running check on the entire assembly chain; measured values are between 1.1×10^{-16} and 3.8×10^{-16} (§S14).

S2.3 Normalization

Let $\widetilde{\omega}$ be a simple root. Differentiating (S3) with respect to any parameter p and contracting with $\boldsymbol{v}$ gives

$$\boldsymbol{v}^T(\partial_p \mathbf{A})\boldsymbol{u} + \frac{\mathrm{d}\widetilde{\omega}}{\mathrm{d}p}\, \boldsymbol{v}^T(\partial_\omega \mathbf{A})\boldsymbol{u} = 0, \tag{S4}$$

where the field derivative $\partial_p \boldsymbol{u}$ has dropped out because $\boldsymbol{v}^T\mathbf{A} = 0$. The quantity

$$\mathcal{M} = \boldsymbol{v}^T(\partial_\omega \mathbf{A})\boldsymbol{u} \tag{S5}$$

is the QNM normalization *for this operator*. It is nonzero exactly when the root is simple, it carries the frequency dependence of the reference Green tensor and the contour, and it is invariant under independent rescaling of $\boldsymbol{v}$ and $\boldsymbol{u}$ in the combination that appears in (S4). Everything in §S12 follows from (S4) alone.

Substituting a closed-cavity energy normalization $\int \epsilon|\mathbf{E}|^2$ for $\mathcal{M}$ is the single most common way to get a plausible-looking $\partial\omega_r/\partial p$ together with a wrong $\partial\omega_i/\partial p$. Since Q depends on ω_i and $|\omega_i|/\omega_r = 1/(2Q)$, that error is invisible in the wavelength and fatal in the objective.

S2.4 The logarithmic objective

Writing $\widetilde{\omega} = a + ib$ with $b < 0$, (S2) gives

$$\delta Q = -\frac{\delta a}{2b} + \frac{a\,\delta b}{2b^2}, \qquad \delta \log Q = \frac{\delta a}{a} - \frac{\delta b}{b}. \tag{S6}$$

We optimize $\log Q$: multiplicative gains become additive, the two components of the pole enter symmetrically in relative terms, and the conditioning of the problem is exposed rather than hidden. Since $|b|/a = 1/(2Q)$, at $Q = 10^4$ the loss rate is 5×10^{-5} of the resonant frequency. Any part of the construction whose relative accuracy is worse than that—an unresolved continuum quadrature, a contour that has drifted onto another sheet, an absorber reflection—can dominate δb and reverse the sign of the gradient while leaving δa apparently converged. This single inequality motivates most of the numerical care documented in §§S7, S10, and S14.

S3 Exact parameterized dielectric and its transforms

S3.1 Reference and contrast

Let the unpatterned device define the reference permittivity ϵ_0. For the air-clad beam this is the rectangular core in a homogeneous cladding,

$$\epsilon_0(y, z) = \epsilon_c + (\epsilon_b - \epsilon_c)\, \mathbf{1}_{|y|<W/2}\, \mathbf{1}_{|z|<T/2}, \tag{S7}$$

and for the multilayer device it is the planar stack with the unpatterned Si_3N_4 strip embedded in it (§S6). In both cases the patterned device differs from the reference only inside the etched holes.

Hole j is the elliptical through-cylinder

$$H_j(\boldsymbol{\theta}) = \left\{\mathbf{r} : \frac{(x - x_j)^2}{r_{x,j}^2} + \frac{y^2}{r_{y,j}^2} \le 1,\ |z| \le \frac{T}{2}\right\}, \tag{S8}$$

so that

$$\epsilon_{\boldsymbol{\theta}} = \epsilon_0 + \Delta\epsilon_{\boldsymbol{\theta}}, \qquad \Delta\epsilon_{\boldsymbol{\theta}} = (\epsilon_c - \epsilon_b) \sum_{j=1}^{N_h} \mathbf{1}_{H_j(\boldsymbol{\theta})}. \tag{S9}$$

For a mirror-symmetric cavity only the positive-x half is stored,

$$x_1 = \tfrac{s_0}{2}, \qquad x_j = \tfrac{s_0}{2} + \sum_{\ell=1}^{j-1} s_\ell, \tag{S10}$$

with holes at $-x_j$ generated exactly. The free vector is therefore

$$\boldsymbol{\theta} = \big(s_0, \ldots, s_{N_h-1};\ r_{x,1}, \ldots, r_{x,N_h};\ r_{y,1}, \ldots, r_{y,N_h}\big), \tag{S11}$$

of length $3N_h$: 45 coordinates for $N_h = 15$ and 126 for $N_h = 42$. Storing pitches rather than absolute centres makes the mirror-side extent a derived quantity, and is why the optimizer's accumulated pitch reduction on the second device shows up as a 268 nm inward motion of the outermost hole (§S17).

S3.2 Closed-form Fourier factors

With $\widehat{f}(\mathbf{k}) = \int f(\mathbf{r}) e^{-\mathrm{i}\mathbf{k}\cdot\mathbf{r}}\, \mathrm{d}^3 r$, direct integration of (S8) over the elliptical disk and the vertical interval gives, exactly,

$$\widehat{\mathbf{1}}_{H_j}(\mathbf{k}) = e^{-\mathrm{i}k_x x_j} \underbrace{\left[2\pi r_{x,j} r_{y,j} \frac{J_1(\rho_j)}{\rho_j}\right]}_{\text{elliptical disk}} \underbrace{\left[\frac{2\sin(k_z T/2)}{k_z}\right]}_{\text{thickness}}, \tag{S12}$$

$$\rho_j^2 = (k_x r_{x,j})^2 + (k_y r_{y,j})^2. \tag{S13}$$

Both bracketed factors have removable singularities, with limits $J_1(\rho)/\rho \to \frac{1}{2}$ and $2\sin(k_z T/2)/k_z \to T$, so that $\widehat{\mathbf{1}}_{H_j}(\mathbf{0}) = \pi r_{x,j} r_{y,j} T$ is exactly the hole volume. The implementation evaluates both through small-argument series below a fixed switchover, which is tested against the closed forms to machine precision. Summing (S12) over holes gives the complete geometry dependence of the scatterer,

$$\widehat{\Delta\epsilon_{\boldsymbol{\theta}}}(\mathbf{k}) = (\epsilon_c - \epsilon_b) \sum_j e^{-\mathrm{i}k_x x_j} \Big[2\pi r_{x,j} r_{y,j} \tfrac{J_1(\rho_j)}{\rho_j}\Big] \Big[\tfrac{2\sin(k_z T/2)}{k_z}\Big]. \tag{S14}$$

This is an exact analytical representation of the entire finite hole array as a function of every centre and every radius. No permittivity is ever sampled onto a grid, and no subpixel smoothing rule enters the model.

S3.3 Analytic geometry derivatives

The derivatives needed by §S12 follow from (S12) in closed form. Translation is a phase,

$$\frac{\partial \widehat{\mathbf{1}}_{H_j}}{\partial x_j} = -\mathrm{i}k_x\, \widehat{\mathbf{1}}_{H_j}, \tag{S15}$$

and for the radii, writing $f(\rho) = J_1(\rho)/\rho$ with

$$f'(\rho) = \frac{J_0(\rho)}{\rho} - \frac{2J_1(\rho)}{\rho^2}, \qquad \frac{\partial \rho_j}{\partial r_{y,j}} = \frac{k_y^2 r_{y,j}}{\rho_j}, \tag{S16}$$

the product rule applied to $2\pi r_{x,j} r_{y,j} f(\rho_j)$ gives

$$\frac{\partial \widehat{\mathbf{1}}_{H_j}}{\partial r_{y,j}} = \widehat{\mathbf{1}}_{H_j} \left[\frac{1}{r_{y,j}} + \frac{f'(\rho_j)}{f(\rho_j)} \frac{k_y^2 r_{y,j}}{\rho_j}\right], \tag{S17}$$

and symmetrically for $r_{x,j}$ with $k_y \to k_x$. Because (S10) is affine, the derivative with respect to a pitch s_ℓ is the sum of (S15) over every hole outboard of ℓ; this is the only place where the geometry parameterization couples coordinates, and it is handled by the reverse-mode accumulation of §S12 rather than by an explicit chain of sums.

For completeness, the rectangular reference contrast of (S7) has the exact transverse transform

$$\widehat{\Delta\epsilon}_{\text{beam}}(k_y, k_z) = (\epsilon_b - \epsilon_c)\, WT\, \operatorname{sinc}\!\left(\tfrac{k_y W}{2}\right) \operatorname{sinc}\!\left(\tfrac{k_z T}{2}\right), \tag{S18}$$

with unnormalized $\operatorname{sinc}(x) = \sin(x)/x$; the beam width W and thickness T are therefore differentiable in the same framework, though both were held fixed in the campaigns reported here.

S3.4 What this parameterization buys

Three properties matter downstream. The model is exact in the geometry variables, so a gradient is a derivative of the true parameterized problem and not of a discretization of it. The derivatives are analytic, so the 6.71×10^{-6}-level agreement with central differences reported in §S14 tests the *implementation* rather than a step-size choice. And the geometry map is continuous as the step size vanishes—no remeshing, no polygon repair, no signed-distance reinitialization—so a line search satisfies $\operatorname{dist}(\Gamma(\alpha), \Gamma(0)) \to 0$ as $\alpha \to 0$ by construction. That last property is not free in general and its violation is documented in §S18.

S4 Outgoing integral equation and the regularized determinant

S4.1 Reduction to the holes

Let $\mathcal{L}_0(\omega) = \nabla \times \nabla \times -(\omega^2/c^2)\epsilon_0$ be the reference Maxwell operator and let $\mathbf{G}^+_{\text{wg}}(\omega) = \mathcal{L}_0(\omega)^{-1}$ be its *outgoing* inverse, constructed in §S5–§S7. Splitting (S1) as $\mathcal{L}_0\mathbf{E} = (\omega^2/c^2)\Delta\epsilon_{\boldsymbol{\theta}}\mathbf{E}$ and inverting gives the homogeneous Lippmann–Schwinger equation

$$\mathbf{E} = \frac{\omega^2}{c^2}\, \mathbf{G}^+_{\text{wg}}(\omega)\, \Delta\epsilon_{\boldsymbol{\theta}}\, \mathbf{E}. \tag{S19}$$

Because $\Delta\epsilon_{\boldsymbol{\theta}}$ vanishes outside $\Omega = \bigcup_j H_j$, the unknown in (S19) may be restricted to Ω: the field everywhere else is recovered by one application of $\mathbf{G}^+_{\text{wg}}$. This is the entire reason the method is cheap. The support of the unknown is a few dozen small ellipses, not a computational domain, and it does not grow when the cavity is made longer or the mirror stronger.

Defining $K(\omega, \boldsymbol{\theta}) = (\omega^2/c^2)\mathbf{G}^+_{\text{wg}}(\omega)\Delta\epsilon_{\boldsymbol{\theta}}$ restricted to Ω, a nontrivial solution of (S19) exists exactly when $I - K$ has nontrivial kernel, i.e. when

$$D(\omega, \boldsymbol{\theta}) = \det\nolimits_{\text{reg}}\big[I - K(\omega, \boldsymbol{\theta})\big] = 0. \tag{S20}$$

Together with the passivity requirement $\operatorname{Im}\widetilde{\omega} < 0$ and the outgoing choice of sheet, (S20) is an exact implicit definition of $Q(\boldsymbol{\theta})$:

$$Q(\boldsymbol{\theta}) = -\frac{\operatorname{Re}\widetilde{\omega}(\boldsymbol{\theta})}{2\,\operatorname{Im}\widetilde{\omega}(\boldsymbol{\theta})}, \quad D(\widetilde{\omega}(\boldsymbol{\theta}), \boldsymbol{\theta}) = 0, \quad \operatorname{Im}\widetilde{\omega} < 0. \tag{S21}$$

S4.2 Why the determinant must be regularized

The subscript in (S20) is not decorative. The electric dyadic Green tensor contains a singular local term—in momentum space the longitudinal projector $-P_L/k_c^2$ of (S27), in real space a δ-function plus a nonintegrable $1/R^3$ tail whose principal value depends on the shape of the excluded volume [6]. The operator K is consequently *not* trace class, and the Fredholm determinant $\det(I - K)$ does not exist as a convergent product of eigenvalues.

Three consequences follow.

1. A collocation matrix built by sampling the real-space kernel at quadrature nodes has entries that depend on the exclusion rule chosen at coincident points. Its determinant converges to *something* under mesh refinement, but not to (S20), and calling it exact would be wrong. We do not use one.
2. We keep the longitudinal projector in momentum space, where it is algebraic. The self term is then never evaluated as a singular real-space integral and no depolarization factor is guessed.
3. The determinant is taken in the regularized sense: either as an ordinary Fredholm determinant of the trace-class remainder after the self term is extracted, or as the two-regularized determinant

$$\det{}_2(I-K) = \det\bigl[(I-K)e^{K}\bigr], \tag{S22}$$

which is well defined for Hilbert–Schmidt K [7]. Both have the same zero set as (S20), since e^{K} is invertible; only the value away from the root differs.

Formally, where the local trace expansions converge,

$$\det(I-K) = \exp\Bigl[-\sum_{m\geq 1}\frac{\operatorname{Tr}K^m}{m}\Bigr], \quad \det{}_2(I-K) = \exp\Bigl[-\sum_{m\geq 2}\frac{\operatorname{Tr}K^m}{m}\Bigr], \tag{S23}$$

the $m=1$ term being exactly the divergent self trace that (S22) removes. Equation (S23) identifies the analytic object; it is not how we evaluate it. At a resonance an eigenvalue of K approaches unity and the trace-log series loses its convergence, so the computational realization is always the finite matrix determinant of §S9.

S4.3 Finite sequence

Projecting (S19) onto the truncated interior basis of §S8 gives a finite matrix

$$A^{(N)}_{ab}(\omega,\boldsymbol{\theta}) = \delta_{ab} - \frac{\omega^2}{c^2}(\epsilon_c-\epsilon_b)\int_\Omega d^3r\int_\Omega d^3r'\,\boldsymbol{\phi}_a\cdot\mathbf{G}^{+}_{\mathrm{wg}}\cdot\boldsymbol{\phi}_b, \tag{S24}$$

whose determinant $D_N(\omega,\boldsymbol{\theta}) = \det A^{(N)}$ has roots $\widetilde{\omega}_N$ and finite quality factors $Q_N = -\operatorname{Re}\widetilde{\omega}_N/(2\operatorname{Im}\widetilde{\omega}_N)$. The claim is convergence of the joint limit

$$Q(\boldsymbol{\theta}) = \lim_{N_{xy},N_z,N_\beta\to\infty} Q_{N_{xy},N_z,N_\beta}(\boldsymbol{\theta}), \tag{S25}$$

in which N_β is a *quadrature* order for the radiation continuum, not a count of physical radiation modes. Measured convergence under independent refinement of each index is reported in §S14.

S4.4 Relation to cell scattering matrices

Partitioning the indices of (S24) by hole and eliminating interior blocks by Schur complements produces exactly the cell scattering maps that a transfer-matrix or Redheffer-star treatment would cascade. The difference is the direction of the derivation: here the cell matrices are a stable block factorization of a converged global operator, so the radiation channels they omit are known and controllable, whereas a cascade assembled from independently computed cell matrices carries an unknown truncation of the continuum. That distinction is not academic—it is the same failure mode that made an early revision of this work report a numerically precise and physically meaningless Q (§S18).

S5 Reference Green tensor: homogeneous cladding

S5.1 Fourier reduction along the beam

The unpatterned beam is invariant in x, so the reference Green tensor reduces to a one-dimensional inverse transform,

$$\mathbf{G}_{\mathrm{wg}}(x,\mathbf{r}_\perp;x',\mathbf{r}'_\perp;\omega) = \int_{\mathcal{C}_\beta}\frac{d\beta}{2\pi}\,e^{i\beta(x-x')}\,\widetilde{\mathbf{G}_{\mathrm{wg}}}(\beta;\mathbf{r}_\perp,\mathbf{r}'_\perp;\omega). \tag{S26}$$

Everything about openness is carried by the contour $\mathcal{C}_\beta$ (§S7); at fixed β the transverse problem is a compactly supported scattering problem, because only the rectangle $|y| < W/2$, $|z| < T/2$ differs from the cladding.

S5.2 Exact homogeneous symbol

Write $k_c = n_c\omega/c$, $\mathbf{k} = (\beta, q_y, q_z)$ and $k^2 = \mathbf{k}\cdot\mathbf{k}$. For the cladding operator $L_c = \nabla\times\nabla\times -k_c^2 I$, introduce the longitudinal and transverse projectors $P_L = \mathbf{k}\mathbf{k}^T/k^2$ and $P_T = I - P_L$. Then

$$\widetilde{\mathbf{G}}_{\mathrm{c}}(\beta, q_y, q_z; \omega) = \frac{P_T}{k^2 - k_c^2} - \frac{P_L}{k_c^2}, \tag{S27}$$

which is verified directly from $[(k^2 - k_c^2)I - \mathbf{k}\mathbf{k}^T]\widetilde{\mathbf{G}}_{\mathrm{c}} = I$. The second term is the singular local part of §S4: in momentum space it is an algebraic, frequency-dependent constant tensor, and keeping it there is what lets us avoid a real-space principal value entirely. The radiation shell is

$$q_y^2 + q_z^2 + \beta^2 = k_c^2, \tag{S28}$$

which after integration over transverse momentum becomes the branch continuum of §S7. For real $|\beta| > k_c$—the guided region—the denominator in (S27) never vanishes on the real transverse plane, which is why guided modes are ordinary poles rather than resonances.

S5.3 Cross-section basis and Galerkin matrices

On $|y| < W/2$ use normalized Legendre functions

$$\phi_m^{(W)}(y) = \sqrt{\tfrac{2m+1}{W}}\, P_m(2y/W), \tag{S29}$$

whose exact Fourier transform is

$$\Phi_m^{(W)}(q_y) = \sqrt{W(2m+1)}\,(-\mathrm{i})^m\, j_m(q_y W/2) \tag{S30}$$

with j_m a spherical Bessel function [8]. Using the analogous basis in z and three Cartesian components, $\boldsymbol{\phi}_{mns} = \phi_m^{(W)}(y)\phi_n^{(T)}(z)\boldsymbol{e}_s$, which is complete in the vector L^2 of the rectangular cross-section, the cladding Green matrix is the explicit double integral

$$[G_c^{(N)}]_{mns,m'n's'} = \int \frac{\mathrm{d}q_y\,\mathrm{d}q_z}{(2\pi)^2}\, \Phi_m^*\Phi_n^*\,[\widetilde{\mathbf{G}}_{\mathrm{c}}]_{ss'}\,\Phi_{m'}\Phi_{n'}. \tag{S31}$$

Every element is a convergent two-dimensional integral of Bessel and rational functions, with no singularity on the real (q_y, q_z) plane once β lies on the deformed contour.

S5.4 Dyson resummation of the core

The beam is restored by resumming its own susceptibility $\chi_b = \epsilon_b - \epsilon_c$ over the rectangular support,

$$\widetilde{\mathbf{G}}_{\mathrm{wg}} = \widetilde{\mathbf{G}}_{\mathrm{c}} + \widetilde{\mathbf{G}}_{\mathrm{c}}\,k_0^2\chi_b\,\widetilde{\mathbf{G}}_{\mathrm{wg}} \implies \widetilde{\mathbf{G}}_{\mathrm{wg}} = \left[I - k_0^2\chi_b\widetilde{\mathbf{G}}_{\mathrm{c}}\right]^{-1}\widetilde{\mathbf{G}}_{\mathrm{c}}. \tag{S32}$$

The finite sequence is $\mathbf{G}_{\mathrm{wg}}^{(N)} = [I - k_0^2\chi_b G_c^{(N)}]^{-1}G_c^{(N)}$ with $N = (N_y, N_z, N_q)$, in which N_q controls the transverse-momentum quadrature in (S31). No absorber thickness or strength appears anywhere in (S32).

S5.5 Guided poles are not inserted by hand

The guided propagation constants are the zeros of

$$\det A_{\mathrm{wg}}(\beta,\omega)=0,\qquad A_{\mathrm{wg}}=I-k_0^2\chi_b\widetilde{\mathbf{G}_{\mathrm{c}}},\tag{S33}$$

i.e. poles of the same resolvent that carries the radiation continuum. At a simple guided pole β_g with null vectors x_g, y_g,

$$\operatorname*{Res}_{\beta=\beta_g}\widetilde{\mathbf{G}_{\mathrm{wg}}}=\frac{x_g\,y_g^{\dagger}\mathbf{G}_{\mathrm{c}}(\beta_g)}{y_g^{\dagger}\partial_\beta A_{\mathrm{wg}}(\beta_g)\,x_g},\tag{S34}$$

so the discrete term of the real-space Green tensor arrives with its normalization fixed by the pole derivative. This matters: open-waveguide modes have no finite-box norm, and normalizing them by one is a standard route to an inconsistent radiation weight [9]. Deforming the contour then splits (S26) into

$$\mathbf{G}_{\mathrm{wg}}^{+}=\mathrm{i}\sum_{g,+}e^{\mathrm{i}\beta_g|x-x'|}\operatorname*{Res}_{\beta_g}\widetilde{\mathbf{G}_{\mathrm{wg}}}+G_{\mathrm{rad}}^{+}+G_{\mathrm{ev}},\tag{S35}$$

the three terms being guided propagation, the outgoing radiation continuum, and the evanescent remainder.

S5.6 On-shell evaluation inside the light cone

For real β strictly inside the light cone, $\kappa^2=k_c^2-\beta^2>0$ and the transverse radial integral in (S31) meets the shell. Writing $H(q)$ for the basis-projected numerator, the identity used is

$$\int_0^Q\frac{q\,H(q)\,\mathrm{d}q}{q^2-\kappa^2-\mathrm{i}0}=\int_0^Q\frac{q\,[H(q)-H(\kappa)]}{q^2-\kappa^2}\,\mathrm{d}q+\frac{H(\kappa)}{2}\log\frac{Q^2-\kappa^2}{\kappa^2}+\frac{\mathrm{i}\pi}{2}H(\kappa).\tag{S36}$$

The first integrand is regular at $q=\kappa$; the last term is the exact on-shell radiation contribution and changes sign for the incoming boundary value. No small artificial loss is added anywhere. The radiation spectral density

$$\rho_{\mathrm{rad}}(\beta)=\frac{\widetilde{\mathbf{G}_{\mathrm{wg}}}^{+}(\beta)-\widetilde{\mathbf{G}_{\mathrm{wg}}}^{-}(\beta)}{2\pi\mathrm{i}}\tag{S37}$$

must be Hermitian positive semidefinite for a passive real-frequency waveguide, and this is an internal test rather than an assumption.

S6 Reference Green tensor: stratified stack

S6.1 Why the homogeneous reference fails here

The second device is not a beam in one cladding. From bottom to top it is: an SiO_2 substrate half-space, $n=1.45$; a 300 nm layer that contains the patterned Si_3N_4 strip, $n=2.0$, with air elsewhere in that layer; 200 nm anthracene, $n=1.8$; 200 nm PVA, $n=1.5$; and an air half-space. The films are invariant in x and y; only the strip is localized.

Replacing the stack by an effective cladding index deletes the guided and leaky poles of the films. Those poles are precisely the channels through which the cavity loses energy vertically, so an effective-index reference cannot support a Q optimization even if it reproduces the resonance wavelength. The correct reference is therefore the *planar stack*, with the strip treated as a localized susceptibility inside the source layer and restored by the same Dyson step as (S32).

The stack also breaks z reflection. The z-parity sector used for the air-clad device is not closed under the multilayer Maxwell operator. Exact x reflection of the symmetric cavity and y reflection of the cross-section survive, so the symmetry projector of §S9 retains a fixed (x,y) sector while carrying both z parities. This roughly doubles the vertical basis at fixed accuracy and is the main reason the second device is more expensive per pole solve than the first.

S6.2 Generalized Fresnel recursion

At fixed in-plane momentum $k_\parallel^2 = \beta^2 + q_y^2$, the vertical wavenumber in layer j is

$$k_{zj} = \sqrt{(n_j k_0)^2 - k_\parallel^2}, \qquad \operatorname{Im} k_{zj} \geq 0, \tag{S38}$$

the branch being fixed by decay rather than by a principal value. TE and TM Fresnel coefficients are composed recursively through each finite layer: if R_j is the reflection seen beyond a layer of thickness d_j, then

$$R_{j-1} = \frac{r_{j-1,j} + R_j\, e^{2\mathrm{i}k_{zj}d_j}}{1 + r_{j-1,j} R_j\, e^{2\mathrm{i}k_{zj}d_j}}. \tag{S39}$$

The recursion retains the exact film poles—the zeros of the denominator—and is holomorphic in every thickness and index away from the branch points of (S38). This is what makes the whole stack differentiable in principle, even though the campaigns here hold the film stack fixed.

S6.3 Scalar vertical kernels and the Wronskian

Let the source layer occupy $0 < z < d$ and let r_b, r_t be the generalized reflections seen at its bottom and top from inside. The two solutions satisfying the respective outgoing reflection conditions are

$$u_L(z) = e^{-\mathrm{i}k_z z} + r_b e^{\mathrm{i}k_z z}, \quad u_R(z) = e^{\mathrm{i}k_z (z-d)} + r_t e^{-\mathrm{i}k_z (z-d)}, \tag{S40}$$

with Wronskian

$$W_\mathrm{r} = 2\mathrm{i}k_z \big(e^{-\mathrm{i}k_z d} - r_b r_t e^{\mathrm{i}k_z d}\big). \tag{S41}$$

The reciprocal scalar kernel normalized by $(\partial_z^2 + k_z^2)g = -\delta$ is then

$$g(z, z') = -\frac{u_L(z_<)\, u_R(z_>)}{W_\mathrm{r}}, \tag{S42}$$

and the guided and leaky modes of the stack are the zeros of (S41). Two independent tests are used on this block: g must be symmetric under $z \leftrightarrow z'$, and its jump in $\partial_z g$ across $z = z'$ must equal -1 to machine precision.

The full vector dyadic is assembled from the TE and TM scalar kernels in the standard way, projected onto the asymmetric cross-section basis in q_y and z, and used exactly where $\widetilde{\mathbf{G}}_\mathrm{c}$ appears for the homogeneous device. For reference, the planar-film TE pole of this stack has $n_\mathrm{eff} \approx 1.55246$ and the Si_3N_4-strip pole $n_\mathrm{eff} \approx 1.76502$; both are tracked separately, because they behave very differently under the contour deformation of §S7.

S6.4 Adding the strip

With $\mathbf{G}_\mathrm{b}(\beta)$ the projected outgoing Green matrix of the complete planar background, the finite-width Si_3N_4 strip is a compact Dyson problem in exactly the sense of (S32),

$$D_\mathrm{s}(\beta) = I - k_0^2 \Delta\epsilon_\mathrm{core} \mathbf{G}_\mathrm{b}(\beta), \qquad \mathbf{G}_\mathrm{wg}(\beta) = D_\mathrm{s}(\beta)^{-1} \mathbf{G}_\mathrm{b}(\beta). \tag{S43}$$

The physical y-odd block of (S43) contains the E_y-dominant feed mode of the cavity. Its pole in β is not a nuisance: it is the mode the cavity is built out of, and its explicit removal from the longitudinal quadrature is the subject of §S10. Attempting the contour integral of (S26) with that pole left in place is the difference between a usable operator and an unusable one, and this is the single largest structural difference between the two devices in this work.

S7 The outgoing contour and its quadrature

S7.1 Branch topology in scaled momentum

Openness is a statement about which side of the branch cuts and poles the longitudinal integration path passes. Working in the scaled variable

$$u = \frac{\beta}{k_c(\omega)}, \qquad k_c = n_c \frac{\omega}{c}, \tag{S44}$$

pins the radiation branch points at $u = \pm 1$ for every frequency, while the guided poles sit near $u = \pm n_{\text{eff}}/n_c$ and move only slowly. The outgoing Sommerfeld path must pass *below* every positive-u singularity and *above* its negative-u partner; this is the statement that a disturbance at $x > x'$ propagates to the right and one at $x < x'$ to the left.

S7.2 The deformation

We use the smooth, closed-form path

$$u(t) = t - \mathrm{i}h \tanh(t/d)\, \exp\bigl[-(t/L)^8\bigr], \qquad -\infty < t < \infty, \tag{S45}$$

with dimensional path $\beta(t, \omega) = k_c(\omega)\, u(t)$. Three properties are used. The tanh factor gives the required opposite excursions on the two sides of the origin with a single smooth function. The super-Gaussian factor returns the path to the real axis outside $|t| \sim L$, so that the far tails, where the integrand is exponentially small and purely evanescent, are integrated on the real axis where the transverse quadrature is cheapest. And the whole path is independent of ω in the scaled variable, so *holding $u(t)$ fixed while moving ω* is an explicit analytic-continuation homotopy from the causal upper half plane to the resonance.

The depth h and width d are not free: h must exceed the largest imaginary excursion any singularity makes during the continuation, and d must be small enough that the path has separated from $u = \pm 1$ before the branch points are reached. Both are verified per solve by monitoring the distance from the path to every tracked singularity; a crossing appears as a discontinuity in $\widetilde{\omega}$ along the homotopy and is rejected rather than accepted with a changed sheet.

S7.3 Orientation test

The first check applied to any implementation of (S45) is the scalar identity

$$\frac{1}{2\pi} \int_{\mathcal{C}_+} \frac{e^{\mathrm{i}\beta x}}{\beta^2 - a^2}\, \mathrm{d}\beta = \frac{\mathrm{i}e^{\mathrm{i}ax}}{2a}, \qquad x > 0, \tag{S46}$$

which fixes the orientation and the sheet unambiguously and fails loudly for the incoming choice. It is cheap enough to run at every frequency in a convergence study, and it is the test that first exposed the wrong-sheet error described in §S18.

S7.4 Panelized quadrature

The integrand of (S26) for a cavity of length L_{cav} oscillates as $e^{\mathrm{i}\beta L_{\text{cav}}}$, so the required node density grows with cavity length. Two quadrature families are implemented. A tangent map compactifies the whole real axis in one rule and is fast, but it hides the omitted high-momentum tail and its convergence cannot be separated from the polynomial order. A panelized finite-cutoff rule is slower and is what all reported numbers use: panels are laid out with breakpoints clustered at $u = \pm 1$ and at the endpoints of the first guided-mode interval, Gauss–Legendre nodes are clustered towards panel endpoints, and the cutoff $\beta_{\max}/k_0$ is refined independently of the panel width and the node count.

For the second device the accepted rule uses 812 positive-β nodes, a maximum dimensionless panel width of 0.025, four endpoint-clustered nodes per panel, and $\beta_{\max}/k_0 = 5$; halving the node count to 412 changes the imaginary part of the tracked eigenvalue by 0.6 % and Q by 0.65 % (§S14). Agreement obtained from a single tangent-map order is not an error estimate and is not reported as one.

S7.5 Known failure modes

Three ways of getting (S45) wrong were encountered and are recorded here because each produces a plausible number rather than an obvious crash.

1. *Real-axis quadrature at complex frequency.* Substituting $\operatorname{Im}\omega < 0$ into a rule built on the real β axis silently lands on the wrong sheet. The resulting $\widetilde{\omega}$ can even be passive; it is simply not the resonance.
2. *Insufficient depth.* If h is too small, a guided pole crosses the path during the continuation and the residue it should have contributed is lost. The symptom is a Q that jumps discontinuously as a geometry parameter is scanned smoothly.
3. *Unclustered panels.* With breakpoints not placed at $u = \pm 1$, the logarithmic and square-root behaviour at the branch points is integrated by a polynomial rule and converges algebraically instead of spectrally. The symptom is a convergence table that improves like N^{-2} and never settles.

S8 Aperture basis and projection

S8.1 Normalized hole coordinates

Inside hole j introduce

$$x - x_j = r_{x,j}\rho\cos\theta, \quad y = r_{y,j}\rho\sin\theta, \quad 0 \le \rho \le 1, \tag{S47}$$

with $z \in [-T/2, T/2]$ and Jacobian $r_{x,j}r_{y,j}\rho$. The map is affine in the semi-axes, which is what makes the basis functions themselves differentiable in the geometry rather than merely defined on a geometry-dependent domain.

S8.2 Zernike–Legendre vector basis

On the unit disk use the real Zernike functions

$$Z^c_{nm} = R^m_n(\rho)\cos m\theta, \qquad Z^s_{nm} = R^m_n(\rho)\sin m\theta, \tag{S48}$$

for $0 \le m \le n$ with $n - m$ even, the sine family omitted at $m = 0$. Their squared norms on the unit disk are

$$N_{n0} = \frac{\pi}{n+1}, \qquad N_{nm} = \frac{\pi}{2(n+1)}\ (m > 0), \tag{S49}$$

and the radial polynomial is evaluated in the numerically stable Jacobi form

$$R^m_n(\rho) = (-1)^{(n-m)/2}\rho^m P^{(m,0)}_{(n-m)/2}(1 - 2\rho^2), \tag{S50}$$

rather than by the explicit alternating sum, which loses relative accuracy above $n \sim 12$ [10, 8]. Tensoring with the vertical Legendre functions and the three Cartesian unit vectors gives

$$\boldsymbol{\phi}_{j,nm\sigma\ell s} = \frac{Z^\sigma_{nm}(\rho,\theta)}{\sqrt{r_{x,j}r_{y,j}N_{nm}}}\sqrt{\frac{2\ell+1}{T}}\, P_\ell(2z/T)\,\boldsymbol{e}_s, \tag{S51}$$

which is complete in $L^2(H_j)^3$ as the retained degrees grow. At $(n, m, \ell) = (0, 0, 0)$ it reduces to the constant $1/\sqrt{\pi r_{x,j}r_{y,j}T}$, so the lowest order of this basis is algebraically identical to the constant-hole projection used in the earliest revision of the solver—a useful continuity check between development stages.

S8.3 Exact thickness projection

Because the hole spans the complete beam thickness and (S51) uses the same normalized Legendre functions in z as the cross-section basis of (S29),

$$\int_{-T/2}^{T/2} \phi_{i_z}(z)\, \phi_\ell(z)\, \mathrm{d}z = \delta_{i_z \ell}, \tag{S52}$$

so the vertical projection is exact and contributes no quadrature error. This is a consequence of the through-hole geometry: a partially etched hole would couple vertical orders and require its own quadrature.

S8.4 Projection into the cross-section basis

At fixed β the projection is

$$C_{a,j\mu}(\beta) = \int_{H_j} \mathrm{d}^3 r\ \psi_a(y,z)\, e^{-\mathrm{i}\beta x}\, h_{j\mu}(x,y,z), \tag{S53}$$

which after (S52) reduces to the disk integral

$$\frac{\sqrt{r_x r_y}}{\sqrt{N_{nm}}}\, e^{-\mathrm{i}\beta x_j} \int_0^1 \rho\, \mathrm{d}\rho \int_0^{2\pi} \mathrm{d}\theta\ \phi_{i_y}(r_y \rho \sin\theta)\, Z_{nm}^{\sigma}\, e^{-\mathrm{i}\beta r_x \rho \cos\theta}. \tag{S54}$$

The integrand is entire in β, so Gauss–Legendre radial quadrature together with the periodic trapezoidal rule in θ converges spectrally [11]; the angular rule in particular is exponentially accurate because the θ dependence is analytic and periodic. The phase $e^{-\mathrm{i}\beta x_j}$ is factored out before quadrature so that a hole displacement is exactly a multiplicative phase, which is what makes (S15) exact at the level of the assembled matrix and not merely at the level of the continuum model.

For the simplest useful case—the constant mode, $\ell = 0$, $(n,m) = (0,0)$—the x integral can be done first at fixed y, giving

$$\int \mathrm{d}x\, e^{-\mathrm{i}\beta x} h_{j,s} = \frac{e^{-\mathrm{i}\beta x_j}}{\sqrt{V_j}}\, 2a_j(y)\, \mathrm{sinc}[\beta a_j(y)]\, \boldsymbol{e}_s, \tag{S55}$$

with $a_j(y) = r_{x,j}\sqrt{1 - y^2/r_{y,j}^2}$ and $V_j = \pi r_{x,j} r_{y,j} T$; the remaining finite y integral is Gauss–Legendre. This closed form is used as a regression test on the general quadrature.

S8.5 Transpose, not conjugate

The projected outgoing Green matrix is

$$\mathbf{G}_H^{+}(\omega) = \int_{\mathcal{C}_+} \frac{\mathrm{d}\beta}{2\pi}\, C(-\beta)^T\, \widetilde{\mathbf{G}_{\mathrm{wg}}}(\beta,\omega)\, C(\beta), \tag{S56}$$

with the left factor $C(-\beta)^T$ and *not* $C(\beta)^\dagger$. For real basis functions and real frequency the two agree, which is exactly why the substitution is tempting and its consequences delayed: the conjugate is not analytic in ω, so a solver built on it converges perfectly well on the real axis and then continues onto nothing in particular. With the transpose, reciprocity predicts $\mathbf{G}_H^{+} = (\mathbf{G}_H^{+})^T$, which is checked numerically at every assembly (§S14).

S8.6 Production orders

The air-clad device uses a $P2$ hole basis with a single vertical order against a $y3$–$z1$ cross-section basis; the multilayer device carries both z parities as required by §S6. Convergence under independent increase of N_{xy}, N_z, and N_β is reported in §S14; the orders quoted here are the production settings at which every reported analytical number was computed, not the largest ever run.

S9 Symmetry reduction and finite-array assembly

S9.1 The assembled operator

Concatenating the per-hole blocks of (S51) into a single index set and inserting (S56) into (S24) gives the operator whose root is the cavity,

$$\mathbf{A}(\omega, \boldsymbol{\theta}) = I - k_0^2(\epsilon_c - \epsilon_b)\,\mathbf{G}_H^{+}(\omega, \boldsymbol{\theta}), \qquad \Delta\epsilon = \epsilon_c - \epsilon_b < 0. \tag{S57}$$

Its dimension is $3N_h \times N_{\text{Zernike}}(N_{xy}) \times (N_z + 1)$ before symmetry reduction: a few thousand for the devices here, which is why a dense factorization is used throughout and why a pole solve is seconds to minutes rather than hours.

S9.2 Symmetry sectors

Mirror symmetry in x is exact for both devices and is used twice: the positive-half parameterization of (S10) halves the number of free coordinates, and the parity projector halves the operator dimension at fixed accuracy. Transverse (y) parity is likewise exact and selects the polarization sector containing the E_y-dominant mode of interest.

Vertical (z) parity is exact only for the air-clad device. The stack of §S6 breaks it, so the multilayer projector retains a fixed (x, y) sector while carrying both z parities. The practical consequence is that a mode which is cleanly labelled by three parities in the first device carries only two labels in the second, and mode tracking (§S11) correspondingly loses one of its cheapest discriminators.

Symmetry is used as a reduction, never as a filter applied after the fact. A mode that leaves its declared sector during a geometry step is a tracking failure and is reported as one.

S9.3 Finite array, not a periodic cell

The array in (S9) is finite and non-uniform: pitch and both radii vary hole by hole. There is no Bloch phase and no supercell anywhere in the construction, and the mirror region is finite, so its residual transmission is part of the computed Q rather than an assumed boundary condition. This is the structural difference from a periodic guided-mode expansion: the object of interest is the finite device, and the leakage out of its ends is one of the loss channels the operator has to represent, alongside vertical radiation.

The cost of a matrix element between holes i and j is dominated by the contour integral (S56), whose integrand carries the phase $e^{\mathrm{i}\beta(x_i - x_j)}$. Distant pairs therefore oscillate faster and need more nodes; this, not the matrix dimension, is what sets the cost of a long cavity and what forced the 812-node panelized rule for the 42-hole device (§S7).

S9.4 Block structure and cell cascades

Ordering the index set by hole gives $\mathbf{A}$ a natural block structure. Schur complements on interior blocks produce cell scattering maps that can be cascaded by Redheffer star products, which is the classical way to treat such an array. We use that structure only as a factorization of an already-assembled $\mathbf{A}$, never as an independent model: a cascade of separately computed cell matrices carries an uncontrolled truncation of the shared radiation continuum, which is one of the specific errors catalogued in §S18. Stated positively, the cell picture is available here as a computational convenience whose approximation error is known, because it is derived from the global operator rather than assumed prior to it.

S10 Feed-pole extraction and residues

S10.1 The problem

For the multilayer device, the strip Dyson step (S43) introduces a simple pole of $\mathbf{G}_{\text{wg}}(\beta)$ at the propagation constant β_f of the E_y-dominant guided feed mode. The contour of §S7 passes around it correctly in principle, but the integrand of (S56) then contains a first-order pole sitting a short distance from a path that must also resolve interference fringes of scale $1/L_{\mathrm{cav}}$. Numerically this is the worst of both worlds: the quadrature must simultaneously resolve a near-singularity and a long oscillation, and the node count required grows without any corresponding gain in accuracy elsewhere.

The remedy is to remove the pole analytically, integrate only what is smooth, and restore the removed part in closed form.

S10.2 Laurent extraction

Let β_f be a simple zero of $\det D_{\mathrm{s}}$ with normalized right and left null vectors r and l. Then

$$\mathbf{G}_{\mathrm{wg}}(\beta) = \frac{R_+}{\beta - \beta_f} + \text{regular}, \qquad R_+ = \frac{r\,[l^H \mathbf{G}_{\mathrm{b}}(\beta_f)]}{l^H D_{\mathrm{s}}'(\beta_f)\,r}, \tag{S58}$$

the derivative being taken with respect to physical β, so that the residue denominator carries units of inverse length. At the real closed-background pole the operator is Hermitian and $l = r$; away from it the general biorthogonal form is used. Reciprocity fixes the partner residue at $-\beta_f$,

$$R_- = -R_+^T, \tag{S59}$$

which is imposed as a check, not as a definition: the two residues are computed independently and their disagreement is a diagnostic on the whole stratified chain.

The regularized integrand is

$$\mathbf{G}_{\mathrm{wg}}^{\mathrm{reg}}(\beta) = \mathbf{G}_{\mathrm{wg}}(\beta) - \frac{R_+}{\beta - \beta_f} - \frac{R_-}{\beta + \beta_f}, \tag{S60}$$

which is smooth across the whole path and is what the panelized quadrature of §S7 actually integrates.

A symmetric two-sided Laurent estimate of R_+—evaluating $(\beta - \beta_f)\mathbf{G}_{\mathrm{wg}}$ at $\beta_f \pm \delta$ and averaging—cancels the linear term of the regular part and therefore converges quadratically in δ, which is how R_+ is obtained to the accuracy needed without differentiating D_{s} symbolically.

S10.3 Exact real-space restoration

The subtracted Laurent terms are not discarded. Closing the contour analytically returns them as ordered feedthrough propagation in real space,

$$\mathbf{G}_{\mathrm{wg}}^{\mathrm{feed}}(x - x') = \mathrm{i}\,R_+\,e^{\mathrm{i}\beta_f(x-x')}\Theta(x - x') + \mathrm{i}\,R_-\,e^{-\mathrm{i}\beta_f(x-x')}\Theta(x' - x), \tag{S61}$$

with Θ the Heaviside step and the value $\frac{1}{2}$ taken at coincident nodes. The ordering matters: (S61) says that the guided mode propagates rightward from a source to everything on its right and leftward to everything on its left, and applying an unordered symmetric form instead double-counts the guided channel between every pair of holes. In matrix terms, $\mathbf{G}_H^+$ acquires an exact rank-structured contribution assembled from the ordered hole positions, added to the numerically integrated regular part.

S10.4 Magnitude of the effect

The subtraction is not a small correction to a working scheme; without it the scheme does not work. The finite-hole self term of the guided channel alone changes the guided block of the assembled matrix by about 8 %, and the near-pole quadrature error it replaces is larger still and is not controlled by refinement at any node count we could afford. The clearest evidence of its importance is the historical comparison in §S18: on one saved geometry direction, the pre-subtraction operator reported an endpoint ratio of 3.6045899—an apparent 3.6× quality-factor improvement—while independent FDTD reported 1.0007628, and the pole-subtracted operator reported 1.0000101. The correct machinery does not merely improve the number; it changes a false positive into a true negative.

S10.5 What must be differentiated

Because (S60) and (S61) both depend on geometry—the former through $C(\beta)$ and the latter through the hole positions in the ordering and through β_f itself—the reverse-mode pass of §S12 must differentiate both. It differentiates the finite-hole Fourier projections in the smooth real-axis continuum and every ordered real-space strip-pole residue. It does not differentiate the rejected legacy contour operator, which is retained in the record but is not on any computational path used here.

S11 Continuation, pole search, and mode identity

S11.1 Locating the first pole

The search begins on the real frequency axis, where the operator is cheapest and its structure best understood. A singular-value scan of $\mathbf{A}(\omega)$ over a frequency window in the declared symmetry sector locates the local minimum of $\sigma_{\min}$; the corresponding left and right singular vectors then seed a local scalar equation

$$\boldsymbol{v}^T \mathbf{A}(\omega)\, \boldsymbol{u} = 0, \tag{S62}$$

which is solved by a secant or Newton iteration in complex ω with $\boldsymbol{v}$ and $\boldsymbol{u}$ updated at each step. Because (S62) is scalar, its root finding is inexpensive relative to the assembly of $\mathbf{A}$, and the cost of a pole solve is dominated by the contour integrals of (S56): the measured production figure is 398.0 s for the air-clad device.

Convergence is declared on three criteria simultaneously: the scalar residual $|\boldsymbol{v}^T \mathbf{A}\boldsymbol{u}|$ below tolerance; the operator residual $\|\mathbf{A}\boldsymbol{u}\|/\|\boldsymbol{u}\|$ at the solve tolerance, measured 2.29×10^{-10} at the production discretization; and $\operatorname{Im}\widetilde{\omega} < 0$. A root that satisfies the first two and violates the third is a continuation failure, not a result.

S11.2 Continuation off the real axis

The frequency is moved from the causal upper half plane to the resonance along a homotopy with the scaled contour $u(t)$ of (S45) held fixed. Two quantities are monitored along the path: the distance from the path to each tracked singularity, and the continuity of $\widetilde{\omega}$ itself. A singularity crossing produces a step in $\widetilde{\omega}$ of order the residue it carries and is rejected. This is the mechanism by which the wrong-sheet error of §S18 is now caught rather than reported.

S11.3 Mode identity under geometry change

Between optimizer steps the geometry changes and the pole must be re-found. Frequency-nearest tracking is insufficient and is the single most dangerous shortcut available here, because a legitimate small shape step can move ω_r by many of the previous linewidths. Recall from (S6) that $|b|/a = 1/(2Q)$: at $Q = 10^4$ a relative frequency shift of 10^{-4}—trivially achievable with a sub-nanometre displacement—is already two linewidths. What must remain continuous is the mode, not a frequency window centred on the old pole.

Candidates are therefore ranked by, in order:

1. biorthogonal field overlap with the accepted QNM, evaluated on a common representation;
2. localization in the intended cavity region;
3. symmetry and polarization sector;
4. agreement with the first-order complex-pole prediction from (S63);
5. frequency proximity, as a final tie-breaker only.

If no candidate clears the overlap and localization gates, the step is declared *unresolved*. An unresolved step does not update the quasi-Newton history and does not enter any reported trajectory. This rule exists because a silent mode switch makes a correct positive directional derivative appear to produce a negative objective change, which will then poison a curvature update for many subsequent steps.

Two representative overlap figures from the record: a physical-template overlap of 0.92776 between the analytical mode and the independently computed field template for the air-clad baseline, and branch overlaps above 0.999614 across the accepted line search of the multilayer device (§S14).

S11.4 Multiple poles and near-degeneracies

Equation (S63) assumes a simple root. Near a degeneracy the normalization $\mathcal{M}$ of (S5) approaches zero and the derivative diverges, which is physically correct—the modes are exchanging identity—and numerically fatal. The implementation monitors $|\mathcal{M}|$ relative to $\|\partial_\omega \mathbf{A}\|\,\|\boldsymbol{v}\|\,\|\boldsymbol{u}\|$ and flags any step at which that ratio falls below a fixed threshold. No such flag was raised on either campaign reported here, which is consistent with both devices being single-mode in their declared sector over the wavelength window used, but it is a limitation of the present treatment rather than a general guarantee: a strongly multimode cavity would need the block generalization of (S63) and is outside the scope of this work.

S12 Pole and quality-factor derivatives

S12.1 The implicit derivative

From (S4), at a simple pole with null vectors $\boldsymbol{v}, \boldsymbol{u}$,

$$\boxed{\frac{\partial \widetilde{\omega}}{\partial \theta_j} = -\frac{\boldsymbol{v}^T(\partial_{\theta_j}\mathbf{A})\boldsymbol{u}}{\boldsymbol{v}^T(\partial_\omega \mathbf{A})\boldsymbol{u}}} \tag{S63}$$

This is exact for the discrete operator: no linearization of the geometry, no perturbative treatment of the radiation, and no assumption that the mode is weakly perturbed. It requires only that the root be simple and that $\partial_{\theta_j}\mathbf{A}$ be available.

Separating real and imaginary parts of $\widetilde{\omega} = a + \mathrm{i}b$,

$$\frac{\partial Q}{\partial \theta_j} = -\frac{1}{2b}\frac{\partial a}{\partial \theta_j} + \frac{a}{2b^2}\frac{\partial b}{\partial \theta_j}, \tag{S64}$$

and for the objective actually used,

$$\frac{\partial \log Q}{\partial \theta_j} = \frac{1}{a}\frac{\partial a}{\partial \theta_j} - \frac{1}{b}\frac{\partial b}{\partial \theta_j}. \tag{S65}$$

Because $|b|/a = 1/(2Q)$, the second term of (S65) carries the signal and the first is nearly irrelevant: a construction that gets a right to twelve digits and b right to three is useless as a Q optimizer, and this is why the tests of §S14 are stated on the Q derivative rather than on the pole.

S12.2 Assembling $\partial_{\theta_j}\mathbf{A}$

From (S57) the geometry enters $\mathbf{A}$ only through $\mathbf{G}_H^{+}(\omega, \boldsymbol{\theta})$, and from (S56) that dependence is carried entirely by the projection matrices $C(\pm\beta; \boldsymbol{\theta})$—plus, for the multilayer device, by the ordered feed-through term (S61). Differentiating (S56),

$$\partial_{\theta_j}\mathbf{G}_H^{+} = \int_{\mathcal{C}_+} \frac{\mathrm{d}\beta}{2\pi}\Big[\big(\partial_{\theta_j}C(-\beta)\big)^T \widetilde{\mathbf{G}_{\mathrm{wg}}}\, C(\beta) + C(-\beta)^T \widetilde{\mathbf{G}_{\mathrm{wg}}}\,\big(\partial_{\theta_j}C(\beta)\big)\Big] + \partial_{\theta_j}\mathbf{G}_H^{\mathrm{feed}}, \tag{S66}$$

with the contour and $\widetilde{\mathbf{G}_{\mathrm{wg}}}$ independent of geometry. The projection derivatives come from (S15) and (S17): a centre displacement is a multiplicative $-\mathrm{i}\beta$, and a radius derivative is the analytic Bessel expression, both evaluated on exactly the same quadrature nodes as the primal. There is no finite differencing anywhere in the chain and no remeshing.

S12.3 Reverse-mode contraction

Equation (S63) shows that only the scalar $\boldsymbol{v}^T(\partial_{\theta_j}\mathbf{A})\boldsymbol{u}$ is ever needed, not the matrix $\partial_{\theta_j}\mathbf{A}$ itself. Writing $\Lambda = \boldsymbol{u}\boldsymbol{v}^T$, the required scalar is $\mathrm{Tr}[\Lambda^T\partial_{\theta_j}\mathbf{A}]$, and the entire gradient is one reverse pass that seeds Λ at the output and accumulates through the primitives—contour quadrature, Dyson inverse, disk quadrature, Bessel form factors, pitch-to-centre accumulation—in reverse order. The cost is a small constant multiple of one assembly and is independent of the number of geometry coordinates.

Two implementation points are worth recording. First, the pitch parameterization (S10) means the adjoint of the centre-accumulation step is a reverse cumulative sum, so the derivative with respect to an inner pitch automatically collects the contributions of every hole outboard of it. Second, the ordered Heaviside structure of (S61) is differentiated as written, with the $\frac{1}{2}$ at coincident nodes, since the ordering is fixed by hole index and does not change under the small displacements a line search applies.

Measured cost: 370.0 s for the full 45-coordinate gradient of the air-clad device, against 398.0 s for the pole solve it differentiates.

S12.4 Boundary form

Equation (S63) can be rewritten as a boundary integral, which is useful both as an independent implementation and as a conceptual statement about which fields matter. For an interface Γ separating isotropic media 1 and 2, with unit normal $\mathbf{n}$ from 1 to 2 and a positive scalar velocity h moving Γ along $\mathbf{n}$,

$$\boldsymbol{v}^T(\partial_p\mathbf{A})\boldsymbol{u} \ \propto \ \int_\Gamma h\Big[(\epsilon_1-\epsilon_2)\,\mathbf{E}_{L\parallel}\cdot\mathbf{E}_{R\parallel} - (\epsilon_1^{-1}-\epsilon_2^{-1})\,D_{L\perp}D_{R\perp}\Big]\mathrm{d}A, \tag{S67}$$

tangential $\mathbf{E}$ and normal $\mathbf{D}$ appearing because those are the components continuous across a sharp dielectric interface [12, 13]. Magnetic contrast contributes the dual form in $\mathbf{H}$ and $\mathbf{B}$ and is absent here. Writing the bracket as a scalar shape-gradient density $g_\omega(s)$,

$$\delta\widetilde{\omega} = \int_\Gamma g_\omega(s)\,h(s)\,\mathrm{d}A \tag{S68}$$

for any admissible normal velocity—which is why hole radii, hole positions, and beam sidewalls are all differentiable in one framework: each is a piece of the same movable boundary.

The proportionality constant and the overall sign in (S67) depend on the precise definition of $\mathbf{A}$, on the orientation of $\mathbf{n}$, and on whether p is a displacement or a signed-distance value. We fix them by test rather than by argument:

1. translate one planar interface by a known signed amount;
2. compare the predicted complex-frequency shift against a centred difference on identical connectivity;
3. reverse the normal and verify that the gradient sign reverses;
4. repeat for every material pair in the device.

S12.5 Boundary metric

The raw density g_Q is generally too rough to advect a boundary directly. It is mapped through a Sobolev metric

$$(I-\ell^2\partial_s^2)h = g_Q, \tag{S69}$$

with s arclength and ℓ a fabrication-aware smoothing length. This is a change of inner product, not an additional objective term: the ascent direction it produces is still an ascent direction for the same Q. Separate hole and sidewall blocks may use different ℓ, but their RMS normalizations must be declared explicitly so that one block does not dominate merely by carrying more sample points.

The geometry handed to the next solve must be the geometry the motion actually produced. Smoothing, signed-distance reinitialization, polygon repair, and remeshing can each introduce a displacement that does not vanish as the requested step does, and a line search is only meaningful when

$$\operatorname{dist}(\Gamma(\alpha), \Gamma(0)) \to 0 \quad \text{as} \quad \alpha \to 0. \tag{S70}$$

For the parameterization of §S3 this holds by construction. It did not hold for an earlier mesh-based revision, which is documented in §S18.

S12.6 Trust ladder

Not every derivative in a project earns the same claim. We use an explicit five-level ladder and state which level each reported quantity has reached.

level	evidence	permitted claim
algebraic	adjoint agrees with a discrete directional derivative	the implementation differentiates its discrete operator
local surrogate	predicted and solved pole changes agree on common connectivity at high mode overlap	the surrogate has a valid local ascent direction
representation transfer	the saved contour reproduces on an independent mesh	the boundary handoff is not dominating the step
physical transfer	periodic independent-solver evaluation improves under a fixed protocol	the optimization improves a physical observable
converged device	Q stable to mesh, domain, run time, and fit choices	a numerical Q may be quoted as validated

In this work the Q shape derivative has passed the first four levels; the fifth is exactly what the panel of §S15 exists to establish and is why both device gains remain provisional. Mode volume and coupling derivatives have *not* passed their own algebraic and local tests to the same standard and are therefore excluded from the quasi-Newton objective entirely; V_{eff} appears in this work only as a reported diagnostic, never as an optimized quantity.

S13 Optimizer, trust region, and acceptance gates

S13.1 Objective and variables

The optimizer maximizes $\log Q$ over the geometry vector $\boldsymbol{\theta}$ of §S3, with the beam width and the layer thicknesses held fixed. Bound constraints keep every semi-axis positive and prevent adjacent holes from overlapping; they were not active at any accepted step in either campaign.

Quasi-Newton curvature is accumulated by BFGS (air-clad device: a short constrained ascent; multilayer device: a pure BFGS run of 105 accepted steps out of 106 attempted). Curvature is updated only after a clean accepted secant—an unresolved or rejected step contributes nothing—because a single poisoned pair persists in the approximate inverse Hessian for many subsequent iterations.

S13.2 Step control

Each proposed step is scaled so that its largest single-coordinate displacement does not exceed a declared bound. For the air-clad campaign the bound was 1 nm of scaled L^2 displacement per step, which realized a maximum single-coordinate move of about 0.34 nm. This is deliberately conservative: it keeps the step inside the regime where the first-order prediction from (S63) is checkable against the solved pole, which is the local-surrogate level of the trust ladder.

A line search evaluates candidate poles on common connectivity—the same quadrature nodes, the same basis, the same contour—so that the measured change is a property of the geometry and not of a re-discretization. A representative accepted line search from the multilayer device:

step	max coord. change	predicted Q_{ana}	branch overlap
-0.50 nm	-0.50 nm	5 492.77	0.999614
-0.25 nm	-0.25 nm	6 348.88	0.999905
baseline	0	7 270.34	1.000000
$+0.25$ nm	$+0.25$ nm	8 189.34	0.999907
$+0.50$ nm	$+0.50$ nm	8 997.81	0.999632

A $+1$ nm probe predicted a still larger Q but its branch overlap fell below the gate of 0.999 and was excluded. That exclusion is the point of the gate: the larger step was not obviously wrong, it simply could no longer be certified to be the same mode.

S13.3 Acceptance gates

A step is accepted only if all of the following hold.

1. *Pole certified.* Passive sign, residual at tolerance, complex symmetry within the running bound (§S11).
2. *Mode resolved.* Biorthogonal overlap and localization gates cleared, declared symmetry sector retained, branch overlap above 0.999.
3. *Geometry valid.* All bounds satisfied; no hole overlap; the applied displacement equals the requested displacement, i.e. (S70).
4. *Surrogate gain positive.* The solved $\log Q$ at the applied geometry exceeds the previous accepted value. A merely predicted gain is not sufficient.

Failure of any gate makes the step *unresolved* rather than merely rejected: the trust radius shrinks, no curvature is recorded, and the state is logged. Of 106 attempted multilayer iterations, 105 were accepted.

S13.4 One loop, written out

For a single accepted geometry the loop is:

1. solve and certify the QNM (§S11);
2. compute the complex boundary gradient from matched left/right fields (§S12);
3. regularize it in the declared boundary metric (S69);
4. predict $\delta\tilde{\omega}$ and $\delta \log Q$ using the *applied* boundary motion, not the requested one;
5. solve candidate poles on common connectivity for the line search;
6. reject unresolved tracking, invalid geometry, poor transfer, or a nonpositive measured surrogate gain;
7. update the quasi-Newton history only after a clean accepted secant;
8. periodically rebuild the geometry independently and evaluate it with an independent solver (§S15).

Steps 5 and 8 answer different questions and are never substituted for one another. Common connectivity makes the local derivative *measurable*; independent reconstruction tests whether the representation and its discretization *transfer*.

S13.5 What the independent solver is not allowed to do

The independent FDTD layer is a qualification instrument. It supplies no gradient, participates in no line search, and cannot cause a step to be accepted. Its diagnostics are recorded at a fixed cadence—every three accepted steps in the multilayer campaign—and are read after the fact. This separation is what makes the transfer statement of the main text meaningful: if FDTD results had been allowed to influence step selection, agreement between the two would be partly circular.

Evidence

S14 Analytical convergence and gradient validation

This section reports what the analytical construction has been shown to do. It establishes the first two levels of the trust ladder of §S12 and no more: nothing here bears on whether the truncated operator's absolute Q equals a converged Maxwell solution.

S14.1 Operator identities

At every accepted state the following are computed and recorded.

Passivity.
$\operatorname{Im}\widetilde{\omega} < 0$. A violation is a continuation failure and aborts the step.

Residual.
$\|\mathbf{A}\boldsymbol{u}\|/\|\boldsymbol{u}\|$ at the solve tolerance; measured 2.29×10^{-10} for the air-clad production operator.

Null-vector equations.
Maximum left and right equation errors below 2.35×10^{-15}.

Complex symmetry.
$\|\mathbf{A} - \mathbf{A}^T\|/\|\mathbf{A}\|$, predicted zero by reciprocity; measured between 1.1×10^{-16} and 3.8×10^{-16} across all accepted multilayer runs, and 3.05×10^{-14} for the air-clad preflight.

Residue reciprocity.
The scattering-residue partner relation (S59), measured 1.53×10^{-15}.

Contour orientation.
The closed-form identity (S46), which is exact and therefore checked to machine precision.

The air-clad release preflight, run on the packaged production operator, gave a passive pole $\widetilde{\omega}/\omega_0 = 0.9850125798555 + -6.3776704224 \times 10^{-5}\,\mathrm{i}$, $Q_{\text{ana}} = 7\,722.35$, and a physical-template overlap of 0.92776, with the full 45-coordinate reverse-mode gradient completing in 370.0 s after a 398.0 s pole solve.

S14.2 Refinement convergence

The three truncation indices of (S25) are refined independently, because agreement obtained by moving two at once is not an error estimate. For the multilayer device the accepted longitudinal rule uses 812 positive-β nodes with maximum dimensionless panel width 0.025, four endpoint-clustered nodes per panel, and $\beta_{\max}/k_0 = 5$. Halving the node count and doubling the panel width gives:

quantity	coarse	accepted	change
max panel width	0.05	0.025	2× refinement
β nodes	412	812	1.97×
Im tracked eigenvalue	0.0018993	0.00188793	0.6 %
Q_{ana} (directional-gradient run)	7 223.48	7 270.74	0.65 %
cutoff $\beta_{\max}/k_0$	4	5	—

A sub-percent change in the imaginary part under a factor-two refinement of the quantity that generates it is the strongest available statement about this block. It is not a claim of absolute accuracy: the ratios in §S17 show the truncated operator disagreeing with an independent solver by more than an order of magnitude at some geometries, and the two statements are compatible because they concern different things—the internal convergence of a truncation, and its distance from physical reality.

S14.3 Gradient validation

The analytic reverse-mode gradient is compared against central finite differences taken on identical connectivity, with the geometry direction taken from a saved optimizer step and scaled to a 0.1 nm maximum coordinate displacement. For the multilayer device at the development grid, the tracked eigenvalue derivative gave

central finite difference	$-9.9476613141 \times 10^{-5} + 6.169167943 \times 10^{-8}\,\mathrm{i}$
analytic VJP	$-9.9476626041 \times 10^{-5} + 6.1700312856 \times 10^{-8}\,\mathrm{i}$

and the corresponding Q derivatives were 3.0939549116 by central difference against 3.0939341486 by the analytic vector–Jacobian product, a relative error of 6.71×10^{-6}.

The Q derivative is the meaningful test, not the eigenvalue derivative: the imaginary parts above agree to about four digits while the real parts agree to eight, and (S65) weights the former far more heavily. Agreement at 6.71×10^{-6} on the Q derivative is therefore a statement about the part of the calculation that actually drives the optimizer.

S14.4 A directional gate

Beyond a pointwise derivative check, the saved iteration-0-to-30 geometry direction was magnified until its largest coordinate changed by 1 nm and scanned:

direction coordinate	Q_{ana}	branch overlap
−1	7 217.44	0.99999942
0	7 223.48	1.00000000
+1	7 226.87	0.99999943

The actual saved iteration-30 displacement is 1/63.39 of that test step, so the rescaled predicted endpoint ratio is ≈ 1.0000103. That is consistent with the direct checkpoint calculation and, importantly, correctly classifies the direction as flat rather than as an improvement—see §S18, where the pre-repair operator classified the same direction as a 3.6× gain.

S14.5 Complete audited value table

Every value quoted in this section is extracted from the retained gate records by a script that verifies the literal string still appears in the cited file, and fails otherwise.

quantity	value	source record
quan_operator_residual	2.29e-10	VALIDATION_EVIDENCE.md
quan_reciprocity_error	3.05e-14	VALIDATION_EVIDENCE.md
quan_template_overlap	0.92776	VALIDATION_EVIDENCE.md
quan_null_vector_error	2.35e-15	VALIDATION_EVIDENCE.md
quan_residue_reciprocity	1.53e-15	VALIDATION_EVIDENCE.md
quan_pole_solve_seconds	398.0 s	VALIDATION_EVIDENCE.md
quan_gradient_seconds	370.0 s	VALIDATION_EVIDENCE.md
quan_preflight_analytical_q	7722.35	VALIDATION_EVIDENCE.md
quan_unverified_continuation_q	30307.26	VALIDATION_EVIDENCE.md
d23_beta_nodes_coarse	412	SUBTRACTED_REAL_AXIS_QNM_GATE.md
d23_beta_nodes_accepted	812	SUBTRACTED_REAL_AXIS_QNM_GATE.md
d23_panel_width	0.025	SUBTRACTED_REAL_AXIS_QNM_GATE.md
d23_beta_cutoff	5 k0	SUBTRACTED_REAL_AXIS_QNM_GATE.md
d23_im_eigenvalue_coarse	0.0018993	SUBTRACTED_REAL_AXIS_QNM_GATE.md
d23_im_eigenvalue_accepted	0.00188793	SUBTRACTED_REAL_AXIS_QNM_GATE.md
d23_im_eigenvalue_change_percent	0.6 %	SUBTRACTED_REAL_AXIS_QNM_GATE.md
d23_q_coarse	7223.48	SUBTRACTED_REAL_AXIS_QNM_GATE.md
d23_q_accepted	7270.74	SUBTRACTED_REAL_AXIS_QNM_GATE.md
d23_q_change_percent	0.65 %	SUBTRACTED_REAL_AXIS_QNM_GATE.md
d23_reciprocity_min	1.1e-16	SUBTRACTED_REAL_AXIS_QNM_GATE.md
d23_reciprocity_max	3.8e-16	SUBTRACTED_REAL_AXIS_QNM_GATE.md
d23_gate_fdtd_q	6176.55	SUBTRACTED_REAL_AXIS_QNM_GATE.md
d23_gate_analytical_q	7270.34	SUBTRACTED_REAL_AXIS_QNM_GATE.md
d23_grad_relative_error	6.71e-06	REPAIRED_OPTIMIZER_FDTD_GATE.md
d23_grad_fd_q_derivative	3.0939549116	REPAIRED_OPTIMIZER_FDTD_GATE.md
d23_grad_vjp_q_derivative	3.0939341486	REPAIRED_OPTIMIZER_FDTD_GATE.md

quantity	value	source record
d23__flat__direction__fdtd__ratio	1.0007628	SUBTRACTED_REAL_AXIS_QNM_GATE.md
d23__flat__direction__legacy__ratio	3.6045899	SUBTRACTED_REAL_AXIS_QNM_GATE.md
d23__flat__direction__repaired__ratio	1.0000101	SUBTRACTED_REAL_AXIS_QNM_GATE.md
d23__line__search__min__overlap	0.999614	REPAIRED_OPTIMIZER_FDTD_GATE.md
d23__branch__overlap__gate	0.999	REPAIRED_OPTIMIZER_FDTD_GATE.md
quan__pole__real	0.9850125798555 w0	VALIDATION_EVIDENCE.md
d23__n__substrate	1.45	DERIVATION_PHASE1.md
d23__n__strip	2.0	DERIVATION_PHASE1.md
d23__n__anthracene	1.8	DERIVATION_PHASE1.md
d23__n__pva	1.5	DERIVATION_PHASE1.md
d23__t__anthracene	200 nm	DERIVATION_PHASE1.md
d23__t__pva	200 nm	DERIVATION_PHASE1.md
d23__t__source__layer	300 nm	DERIVATION_PHASE1.md
d23__gate__pole__real	0.9992524494 w0	SUBTRACTED_REAL_AXIS_QNM_GATE.md
d23__gate__pole__imag	-6.871739e-05 w0	SUBTRACTED_REAL_AXIS_QNM_GATE.md
d23__gate__lambda__analytical	796.28805 nm	SUBTRACTED_REAL_AXIS_QNM_GATE.md
d23__gate__lambda__fdtd	796.6645 nm	SUBTRACTED_REAL_AXIS_QNM_GATE.md
d23__gate__q__error__percent	17.72 %	SUBTRACTED_REAL_AXIS_QNM_GATE.md
d23__gate__lambda__error__percent	-0.0473 %	SUBTRACTED_REAL_AXIS_QNM_GATE.md
quan__phase57__analytical__q	14568.9	VALIDATION_EVIDENCE.md
quan__phase57__fdtd__baseline	15468.56	VALIDATION_EVIDENCE.md
quan__phase57__fdtd__optimized	42792.17	VALIDATION_EVIDENCE.md
quan__phase57__ratio	2.7664	VALIDATION_EVIDENCE.md
quan__phase58__step1__q	15662.82	VALIDATION_EVIDENCE.md
quan__phase58__step2__q	16818.3	VALIDATION_EVIDENCE.md
quan__phase58__max__coord__nm	0.34 nm	VALIDATION_EVIDENCE.md
quan__pole__imag	-6.3776704224e-05 w0	VALIDATION_EVIDENCE.md
d23__panel__width__coarse	0.05	SUBTRACTED_REAL_AXIS_QNM_GATE.md
d23__node__refinement__factor	1.97 x	SUBTRACTED_REAL_AXIS_QNM_GATE.md
d23__beta__cutoff__coarse	4 k0	SUBTRACTED_REAL_AXIS_QNM_GATE.md
d23__scan__q__minus	7217.44	SUBTRACTED_REAL_AXIS_QNM_GATE.md
d23__scan__q__zero	7223.48	SUBTRACTED_REAL_AXIS_QNM_GATE.md
d23__scan__q__plus	7226.87	SUBTRACTED_REAL_AXIS_QNM_GATE.md
d23__scan__overlap__minus	0.99999942	SUBTRACTED_REAL_AXIS_QNM_GATE.md
d23__scan__overlap__plus	0.99999943	SUBTRACTED_REAL_AXIS_QNM_GATE.md
d23__scan__rescale	63.39	SUBTRACTED_REAL_AXIS_QNM_GATE.md
d23__rescaled__endpoint__ratio	1.0000103	SUBTRACTED_REAL_AXIS_QNM_GATE.md
d23__fd__eigen__real	-9.9476613141e-05	REPAIRED_OPTIMIZER_FDTD_GATE.md
d23__fd__eigen__imag	6.169167943e-08	REPAIRED_OPTIMIZER_FDTD_GATE.md
d23__vjp__eigen__real	-9.9476626041e-05	REPAIRED_OPTIMIZER_FDTD_GATE.md
d23__vjp__eigen__imag	6.1700312856e-08	REPAIRED_OPTIMIZER_FDTD_GATE.md
d23__ls__q__minus__half	5492.77	REPAIRED_OPTIMIZER_FDTD_GATE.md
d23__ls__q__minus__quarter	6348.88	REPAIRED_OPTIMIZER_FDTD_GATE.md
d23__ls__q__baseline	7270.34	REPAIRED_OPTIMIZER_FDTD_GATE.md
d23__ls__q__plus__quarter	8189.34	REPAIRED_OPTIMIZER_FDTD_GATE.md
d23__ls__q__plus__half	8997.81	REPAIRED_OPTIMIZER_FDTD_GATE.md
d23__ls__overlap__minus__half	0.999614	REPAIRED_OPTIMIZER_FDTD_GATE.md
d23__ls__overlap__minus__quarter	0.999905	REPAIRED_OPTIMIZER_FDTD_GATE.md
d23__ls__overlap__plus__quarter	0.999907	REPAIRED_OPTIMIZER_FDTD_GATE.md
d23__ls__overlap__plus__half	0.999632	REPAIRED_OPTIMIZER_FDTD_GATE.md

S15 Independent-FDTD protocol and the ringdown audit

S15.1 Role and separation

The independent solver is a finite-difference time-domain code with perfectly-matched-layer absorbers [14, 15, 16]. It is used only to evaluate frozen geometries. It supplies no gradient, participates in no line search, and no analytical step was ever accepted or rejected on the basis of an FDTD result (§S13). This separation is what makes any observed agreement informative rather than circular.

S15.2 Endpoint procedure

For a comparison between two geometries, both are rendered from the retained geometry vectors through the same generator into simulations that are identical in source spectrum and position, monitor set and position, domain extent, mesh density, absorber configuration, run time, and resonance-analysis code. Only the geometry differs. Quality factors come from an exponential fit to the time-monitor ringdown after the source has decayed, together with a reported fit error; effective mode volume and a localization fraction are recorded alongside.

S15.3 Ringdown adequacy

A ringdown fit constrains the decay rate only if the recorded window is long compared with the amplitude lifetime of the mode. With

$$\gamma = \frac{\omega}{2Q}, \qquad \tau = \frac{1}{\gamma} = \frac{2Q}{\omega}, \qquad \mathcal{N} = \frac{t_{\text{run}} - t_{\text{start}}}{\tau}, \tag{S71}$$

$\mathcal{N}$ counts the lifetimes actually observed. We audited all 35 retained FDTD observations across both campaigns; the complete table appears at the end of this section. 24 of them have $\mathcal{N} < 1$.

Three remarks are needed for this to be read correctly.

1. The air-clad records store a fitted decay rate directly. It reproduces $\omega/(2Q)$ to machine precision, so the two are the same fitted quantity in two forms; the agreement is bookkeeping, not an independent check of the fit.
2. The multilayer diagnostics do not retain a source start time. Their window is therefore bounded above by the total run time and the reported $\mathcal{N}$ is optimistic—the true value is smaller by the source ramp, which for comparable settings on the other device was 1.35 ps.
3. $\mathcal{N} < 1$ does not make a fit meaningless; it makes it poorly conditioned and, in this data set, biased towards larger Q.

The internal evidence for that bias direction is direct. Along the multilayer trajectory, the 12 ps diagnostic at step 61 reports $Q_{\text{FDTD}} = 156\,520$; every subsequent 64 ps analysis, at geometries with strictly higher analytical Q, reports values near 58 240.27. Either the optimizer destroyed a factor of three in physical Q over thirteen steps while its own objective rose, or the short-window fits were overestimating. The protocol change was made at the time for exactly this reason.

S15.4 Predeclared promotion panel

Promotion of a device gain from provisional to validated requires a matched convergence panel that was declared before execution and recorded in machine-readable form. Each endpoint is evaluated over the full factorial of four binary axes:

axis	air-clad	multilayer
steps per wavelength	28 / 35	18 / 20
run time (ps)	60 / 90	64 / 96
PML layers	12 / 16	12 / 16
padding	1.25 / 1.5625	1.0 / 1.25

That is 16 variants per endpoint and 64 tasks in total, with nine required artifacts per task: the immutable simulation description, solver version and task identifier, realized cost, the resonance table with fit diagnostics, mode identity and localization, a geometry plot, a field plot carrying its quantity, slice, normalization and colour scale, the spectrum and ringdown, and artifact checksums.

The promotion gate requires all of: matched endpoint settings; stable mode identity across every variant; passing fit diagnostics; the optimized Q above the initial Q in *every* executed variant; and a central value selected by the predeclared refinement rule. The panel status is predeclared not submitted, and no build command in the source repository can submit it.

S15.5 A required amendment to the panel

The adequacy audit revises the declaration. At the air-clad optimized endpoint's amplitude lifetime of 94.5 ps, the predeclared run times of 60 ps and 90 ps give $\mathcal{N} = 0.63$ and 0.95—still under one lifetime. The declared axis cannot certify a Q of that magnitude. The promotion run must extend the run-time axis until $\mathcal{N} \gtrsim 3$ at the optimized endpoint, which at the historical value implies a window beyond 284 ps. For the multilayer endpoint, whose lifetime is 47.65 ps, the declared 64 ps and 96 ps give $\mathcal{N} = 1.3$ and 2, which is better but still short of that target.

We record this as an amendment rather than reinterpreting the original declaration, because the value of a predeclaration lies entirely in its being harder to change than to satisfy.

S15.6 Complete adequacy table

device	record	Q_{FDTD}	$1/\gamma$ (ps)	window (ps)	$\mathcal{N}$	< 1
quan_loncar	quan_baseline	15 449	24.64	22.65	0.919	•
quan_loncar	quan_optimized_phase58	59 337	94.50	22.65	0.240	•
design23	step_0000	18 992	15.98	12.00	0.751	•
design23	step_0003	18 591	15.59	12.00	0.769	•
design23	step_0006	14 900	12.46	12.00	0.963	•
design23	step_0009	14 307	11.93	12.00	1.006	
design23	step_0012	18 699	15.58	12.00	0.770	•
design23	step_0015	19 826	16.53	12.00	0.726	•
design23	step_0018	20 429	17.02	12.00	0.705	•
design23	step_0021	20 411	17.01	12.00	0.705	•
design23	step_0024	20 458	17.07	12.00	0.703	•
design23	step_0027	22 690	18.94	12.00	0.634	•
design23	step_0030	24 151	20.16	12.00	0.595	•
design23	step_0034	22 262	18.59	12.00	0.646	•
design23	step_0037	17 488	14.57	12.00	0.824	•
design23	step_0040	22 573	18.75	12.00	0.640	•
design23	step_0043	39 625	32.91	12.00	0.365	•
design23	step_0046	55 959	46.46	12.00	0.258	•
design23	step_0049	62 126	51.62	12.00	0.232	•
design23	step_0052	66 985	55.61	12.00	0.216	•
design23	step_0055	58 168	48.29	12.00	0.248	•
design23	step_0058	1.0806×10^5	89.77	12.00	0.134	•
design23	step_0061	1.5652×10^5	130.00	12.00	0.092	•
design23	step_0064	1.3578×10^5	112.73	12.00	0.106	•
design23	step_0067	1.4215×10^5	117.92	12.00	0.102	•
design23	step_0070	436.34	0.38	12.00	31.931	
design23	step_0080	54 658	44.87	64.00	1.426	
design23	step_0083	57 106	46.87	64.00	1.365	
design23	step_0086	56 532	46.39	64.00	1.380	
design23	step_0089	57 304	47.02	64.00	1.361	
design23	step_0092	58 481	47.94	64.00	1.335	
design23	step_0096	54 257	44.46	64.00	1.439	
design23	step_0099	55 039	45.08	64.00	1.420	
design23	step_0102	64 176	52.54	64.00	1.218	
design23	step_0105	58 240	47.65	64.00	1.343	

S16 Air-clad Quan–Lončar campaign

S16.1 Device and parameterization

A mirror-symmetric silicon nanobeam, cross-section 700 nm×220 nm, core index 3.46, suspended in air, with a 990 nm unpatterned lead at each end and 15 elliptical through-holes per side. The free vector is the 15 positive-side pitches and the two semi-axes of each hole: 45 coordinates. The baseline is the deterministic design of Quan and Lončar [17], with a constant 330 nm pitch and radii tapering from 121.27 nm at the cavity centre to 85.75 nm at the mirror; the target wavelength is near 1 502.32 nm in the E_y-dominant sector.

The production analytical model uses the homogeneous air-clad reference of §S5 with a $P2$ hole basis at one vertical order against a $y3$–$z1$ cross-section basis, and the custom reverse-mode gradient of §S12.

S16.2 A lineage caution

The analytical solver for this device was developed against a *different* benchmark: the encapsulated silicon-nitride nanobeam of Fryett *et al.* [18], with $W = 450\,\mathrm{nm}$, $T = 330\,\mathrm{nm}$, $n_b = 2.0$, $n_c = 1.47$, mirror pitch $a_m = 233\,\mathrm{nm}$, and a reported $Q \simeq 10^5$ near 740 nm. The derivation record therefore quotes those material and geometric values throughout its early phases. They are *not* the production device. Any number in this document that describes the campaign refers to the air-clad silicon device above; where a derivation is cited for its mathematics, the mathematics is device-independent and the constants are not carried over.

That benchmark also contains a geometry ambiguity worth recording, since it affects anyone attempting the same reconstruction: the source describes a linear taper of major diameter and period, mirror diameters of 300 nm/100 nm, an innermost major diameter of 100 nm with central holes "separated by 140 nm", and a Bragg region of 40 holes. Those statements do not determine a unique geometry—in particular whether 140 nm is a separation or an optimized transverse diameter, and whether 40 is per side or total. The derivation treats the geometry as a parameter vector and retains the earlier reconstruction only as a named candidate.

S16.3 Optimization

The accepted design is step 3 of a constrained ascent, each step a scaled L^2 displacement of at most 1 nm with a realized maximum single-coordinate move of about 0.34 nm. The analytical quality factor along those steps was $14\,568.9 \to 15\,662.82 \to 16\,818.3 \to 18\,037.3$.

A further eight-step continuation of the same ascent reached $Q_{\mathrm{ana}} = 30\,307.26$ but was never evaluated in FDTD. It is retained in the record under a distinct name so that it cannot be confused with the qualified step-3 design, and it supports no claim here: it is evidence of analytical continuation, not of physical improvement.

S16.4 Geometry outcome

The endpoint geometries reduce to the following descriptors, computed from the retained seed files.

	before		after	
quantity	mean	range	mean	range
pitch (nm)	330	330–330	331.59	328–333.7
r_x (nm)	109.74	85.75–121.3	108.95	86.02–122.9
r_y (nm)	109.74	85.75–121.3	109.87	86.57–121.9
r_y/r_x	1	1–1	1.0086	0.9916–1.016
$2r_y/W$	0.3135	0.245–0.3465	0.3139	0.2473–0.3482
$1 - 2r_x/p$	0.3349	0.265–0.4803	0.3427	0.252–0.4796
area (nm^2)	38 225	2.31e+04–4.62e+04	37 986	2.339e+04–4.705e+04

The reading is unambiguous. The mean in-plane aspect ratio moves from 1 to 1.0086; the largest transverse semi-axis change over all 15 holes is 0.819 nm; the mean hole area changes by a factor 0.9937; and the transverse air fraction is essentially unchanged. The only substantial motion is longitudinal, up to 24.6 nm, outward at the mirror end and inward near the centre, i.e. a mild stretch of the taper envelope. This is local refinement of an existing design, and the value of including it here is precisely that it is *not* a dramatic restructuring: it establishes that the method also works in the regime where the answer is already nearly optimal.

S16.5 Independent evaluation

Matched 24 ps FDTD analyses of the two frozen geometries give:

	baseline	Phase-58 step 3
Q_{FDTD}	15 449.47	59 337.16
λ (nm)	1 502.32	1 499.97
Q_{ana}	7 455.16	18 037.3
$Q_{\text{ana}}/Q_{\text{FDTD}}$	0.483	0.304
$\tau = 1/\gamma$ (ps)	24.64	94.5
window (ps)	22.65	22.65
$\mathcal{N}$	0.919	0.24

The FDTD gain is 3.8407× and the analytical gain is 2.42×: the surrogate under-predicts both the absolute value and the improvement. A slightly earlier revision of the same comparison, at 12 ps rather than 24 ps, gave 15 468.56 → 42 792.17, a ratio of 2.7664; the two ratios differ by 39 % under nothing but a change of ringdown length, which is itself a useful measure of how much the quoted number depends on the protocol.

S16.6 Status

Provisional. The ratio is a matched comparison, but $\mathcal{N} = 0.24$ at the optimized endpoint means the absolute values are not converged, and the predeclared panel of §S15 has not been executed. The amendment recorded there—extending the run-time axis until $\mathcal{N} \gtrsim 3$—applies to this device in particular.

S17 Multilayer Design23 campaign

S17.1 Device and parameterization

The stack, bottom to top: SiO_2 substrate half-space, $n = 1.45$; a 300 nm layer containing the patterned Si_3N_4 strip, $n = 2.0$, air elsewhere in that layer; 200 nm anthracene, $n = 1.8$; 200 nm PVA, $n = 1.5$; air half-space. The strip is 556.264 nm wide and carries 42 elliptical through-holes per side, giving 126 free coordinates: 42 pitches and two semi-axes each. The strip width and the layer thicknesses were held fixed. The resonance is near 792.347 nm in the E_y-dominant sector.

The analytical model uses the stratified reference of §S6, the strip Dyson step (S43), the feed-pole subtraction of §S10, and the 812-node panelized contour rule of §S7. Both z parities are carried because the stack breaks vertical reflection.

A superseded revision of this device used 25 holes per side and 75 coordinates; several derivation phases in the retained record describe that version. The production device is the 42-hole, 126-coordinate one, and all numbers here refer to it.

S17.2 Trajectory

A pure BFGS ascent on $\log Q$ accepted 105 of 106 attempted steps, raising Q_{ana} from 8 958.74 to 717 003, a factor of 80.03. Independent FDTD diagnostics were taken every three accepted steps. The complete retained record follows; $\mathcal{N}$ is the ringdown adequacy of (S71) and † marks the single observation excluded by the predeclared fit-quality rule.

step	Q_{ana}	Q_{FDTD}	λ (nm)	V_{eff} $(\lambda/n)^3$	loc.	fit err.	t_{run} (ps)	$\mathcal{N}$
0	8 958.7	18 992	792.35	5.047	0.679	0.000429	12	0.75
3	12 488	18 591	790.03	4.736	0.691	0.000663	12	0.77
6	14 350	14 900	787.81	4.424	0.705	0.00121	12	0.96
9	15 027	14 307	785.35	4.264	0.700	0.000755	12	1.01
12	16 012	18 699	784.56	4.208	0.693	0.00153	12	0.77
15	17 410	19 826	785.26	4.114	0.695	0.00145	12	0.73
18	18 439	20 429	784.56	4.384	0.687	0.00136	12	0.71
21	19 742	20 411	785.09	4.992	0.688	0.00134	12	0.71
24	21 154	20 458	785.67	4.023	0.690	0.00133	12	0.70
27	22 698	22 690	786.12	4.164	0.689	0.000539	12	0.63
30	24 371	24 151	786.35	4.008	0.687	0.00175	12	0.60
34	30 774	22 262	786.29	4.008	0.695	0.000282	12	0.65
37	36 042	17 488	784.78	3.439	0.690	0.00167	12	0.82
40	49 688	22 573	782.13	3.275	0.690	0.00233	12	0.64
43	63 495	39 625	782.25	3.199	0.691	0.000822	12	0.36
46	85 898	55 959	781.94	3.032	0.686	0.00181	12	0.26
49	1.2595×10^5	62 126	782.55	3.412	0.670	0.000222	12	0.23
52	1.5413×10^5	66 985	781.92	3.133	0.668	0.00113	12	0.22
55	1.7445×10^5	58 168	781.92	3.234	0.669	0.00104	12	0.25
58	2.1063×10^5	1.0806×10^5	782.41	3.228	0.669	0.000441	12	0.13
61	2.3609×10^5	1.5652×10^5	782.25	3.100	0.672	0.000759	12	0.09
64	2.5333×10^5	1.3578×10^5	781.94	3.100	0.672	0.00215	12	0.11
67	2.8578×10^5	1.4215×10^5	781.25	2.876	0.667	0.00223	12	0.10
70 †	3.2658×10^5	436.34	811.16	2.532	0.696	0.0725	12	31.93
80	6.4448×10^5	54 658	773.12	2.690	0.669	0.00675	64	1.43
83	6.5083×10^5	57 106	773.07	2.690	0.669	0.000832	64	1.37
86	6.5993×10^5	56 532	772.92	2.690	0.670	5.22×10^{-6}	64	1.38
89	6.7172×10^5	57 304	772.76	2.689	0.670	3.35×10^{-6}	64	1.36
92	6.7909×10^5	58 481	772.10	2.691	0.671	5.57×10^{-6}	64	1.33
96	7.0438×10^5	54 257	771.77	2.687	0.671	3.62×10^{-6}	64	1.44
99	7.0822×10^5	55 039	771.39	2.687	0.672	9.5×10^{-7}	64	1.42
102	7.0963×10^5	64 176	771.09	2.818	0.672	0.00703	64	1.22
105	7.17×10^5	58 240	770.50	2.652	0.675	1.19×10^{-6}	64	1.34

S17.3 The excluded observation

The step-70 diagnostic reports $Q_{\mathrm{FDTD}} = 436.3$ with a resonance-fit error of 0.0725—about two orders of magnitude above the surrounding scatter—together with a wavelength excursion to 811.16 nm, some 30 nm away from every neighbouring point. It is excluded from all aggregate statistics by the rule declared before the audit was run: resonance error above 0.01. That threshold excludes exactly this one point of 33; the next largest fit error in the record is 0.00703.

The exclusion is a fit-quality judgement, not a value judgement: it was applied without reference to the resulting Q, and the point is retained and marked in every figure rather than deleted. The most likely mechanism is a mode identification failure in the FDTD post-processing—the wavelength excursion and the fit error are consistent with the analysis latching onto a different resonance in the window—but we did not re-run it and do not claim to know.

S17.4 Protocol change

Diagnostics through step 67 used a 12 ps run time; those from step 80 onward used 64 ps. The change was made because the lengthening photon lifetime had overtaken the window: at step 61, $Q_{\mathrm{FDTD}} = 156\,520$ implies $\tau \approx 130$ ps against a 12 ps window, i.e. $\mathcal{N} \approx 0.092$. The two segments are therefore different observables and every aggregate in the main text is reported for each separately as well as together.

S17.5 Geometry outcome

	before		after	
quantity	mean	range	mean	range
pitch (nm)	234.63	234.1–234.9	228.46	202.4–249.5
r_x (nm)	76.157	68.23–90.35	45.96	36.73–78.23
r_y (nm)	76.166	68.25–90.49	118.16	64.92–165.1
r_y/r_x	1.0001	0.9981–1.002	2.5613	1.36–3.112
$2r_y/W$	0.2738	0.2454–0.3253	0.4248	0.2334–0.5938
$1 - 2r_x/p$	0.3508	0.2301–0.4187	0.5986	0.373–0.6791
area (nm^2)	18 464	1.463e+04–2.568e+04	17 542	8743–4.058e+04

The mean longitudinal semi-axis contracts from 76.157 nm to 45.96 nm; the mean transverse semi-axis expands from 76.166 nm to 118.16 nm, with a maximum of 165.1 nm. The mean aspect ratio moves from 1.0001 to 2.5613, maximum 3.112. Near-circular holes become transverse slots spanning up to 0.5938 of the strip width, leaving dielectric rails of 113 nm per side against 187.6 nm initially. The mean hole *area* changes by a factor of only 0.9501: the air was reshaped, not added. The mean pitch contracts from 234.63 nm to 228.46 nm, and because pitches accumulate ((S10)) the outermost hole centre moves inward by 268 nm.

S17.6 Spectral and modal signature

The restructuring is accompanied by a blueshift of 21.8 nm, from 792.347 nm to 770.497 nm, and by a reduction of the effective mode volume from 5.047 to 2.652 $(\lambda/n)^3$, a factor of 0.525. The reported localization fraction is essentially unchanged, 0.679 to 0.675, so the mode did not migrate out of the cavity region.

Concentrating field energy into the air slots at fixed period would produce exactly this combination, and the geometry change is what one would design in order to obtain an air-type rather than a dielectric-type mode. We report it as a *characterized geometric and spectral restructuring consistent with* a change of mode character, and not as a demonstrated band-edge swap. Two reasons for that caution:

1. The available evidence is a correlation among hole shape, wavelength, and mode volume. The diagnostics that would settle the question—a band structure of the terminal mirror cell before and after, or a decomposition of the retained field energy between the dielectric and air regions—were not computed, and computing them would require re-analysis of raw field data that is outside the scope of this work.
2. A simple arithmetic argument points the other way. Evaluated on the mean pitch, $\lambda/2p$ is 1.689 at step 0 and 1.686 at step 105. The ratio of wavelength to period is essentially unchanged, so the blueshift is accounted for by the pitch contraction alone and cannot by itself be read as a jump to the other side of a gap.

S17.7 Independent evaluation and status

The exact retained endpoints:

	step 0	step 105
Q_{FDTD}	18 991.94	58 240.27
λ (nm)	792.347	770.497
Q_{ana}	8 958.74	717 003
$Q_{\mathrm{ana}}/Q_{\mathrm{FDTD}}$	0.472	12.31
V_{eff} $(\lambda/n)^3$	5.047	2.652
run time (ps)	12	64
$\mathcal{N}$	0.751	1.34

giving a gain of 3.0666×. The step-0 value is the endpoint of *this* trajectory; a Q of 20 942.06 from a separately qualified seed lineage exists in the record and must not be substituted for it, since doing so would change the quoted gain without any corresponding physical claim.

PROVISIONAL, for two independent reasons. The endpoints were not analyzed under matched settings—12 ps against 64 ps—and the predeclared panel of §S15 has not been executed. The mismatch runs in the direction that inflates the step-0 value and therefore deflates the quoted gain, but a mismatch in a favourable direction is still a mismatch.

S18 Rejected approximations and failures

A method paper that reports only what worked is not reproducible. This section catalogues the approaches that were tried and rejected, in the order they were encountered, with the symptom that exposed each one. Several produced plausible, converged-looking numbers, which is why they are worth recording.

S18.1 Few-mode and closed bases

Approach. Expand the cavity field in a handful of guided modes of the unpatterned beam, or in the modes of a domain terminated by an absorber, and read the linewidth from the resulting eigenvalue.

Symptom. A numerically precise Q that is essentially arbitrary. The few retained modes reproduce the near field well, so the resonance wavelength converges convincingly, while the tiny continuum coupling that actually determines $\operatorname{Im}\tilde{\omega}$ is absent or replaced by an artefact of the absorber.

Resolution. The full guided-plus-continuum Green tensor of §S5. The imaginary part of the pole is generated by the continuum integral in (S35); no artificial linewidth is assigned to any mode anywhere in the construction.

S18.2 Unregularized collocation determinant

Approach. Sample the real-space dyadic Green kernel at quadrature nodes and take the determinant of $I - K$ directly.

Symptom. A determinant whose value depends on the coincident-point exclusion rule and which converges under refinement to something that is not the Fredholm determinant.

Resolution. Keep the longitudinal projector in momentum space and use a regularized determinant, §S4. The self term is then never evaluated as a singular real-space integral and no depolarization factor is guessed.

S18.3 Real-axis quadrature at complex frequency

Approach. Build the longitudinal quadrature on the real β axis, then substitute $\operatorname{Im}\omega < 0$ to find the pole.

Symptom. A converged root, sometimes even passive, that is not the resonance: the substitution has silently continued onto a different sheet.

Resolution. The frozen scaled contour of (S45), moved along an explicit homotopy with singularity tracking, plus the closed-form orientation test (S46) as a precondition on any contour implementation.

S18.4 Hermitian conjugate in place of transpose

Approach. Use $C(\beta)^\dagger$ as the left factor in (S56), by analogy with a real-frequency energy inner product.

Symptom. Perfect agreement on the real-frequency boundary, then a continuation that does not converge to anything stable, and a loss of the complex symmetry that reciprocity requires.

Resolution. $C(-\beta)^T$, §S8. Complex symmetry of the assembled operator is now a running numerical check rather than an assumption.

S18.5 Layered cavity without feed-pole subtraction

Approach. Integrate (S56) for the multilayer device with the guided feed pole left in the integrand and the contour routed around it.

Symptom. A quality factor of order 80—three orders of magnitude below any plausible value—together with quadrature error that did not respond to node refinement at affordable cost.

Resolution. The Laurent extraction and exact real-space restoration of §S10. The rejected result is retained in the record and is superseded by the accepted pole $\widetilde{\omega}/\omega_0 = 0.9992524494 + -6.871739 \times 10^{-5}\,\mathrm{i}$, giving $Q_{\mathrm{ana}} = 7\,270.74$ at 796.28805 nm against $Q_{\mathrm{FDTD}} = 6\,176.55$ at 796.6645 nm: a Q disagreement of 17.72 % and a wavelength disagreement of −0.0473 %.

S18.6 The flat direction that looked like a 3.6× gain

This is the most consequential failure in the record and the clearest single argument for the operator construction of this work.

Setting. A saved geometry direction from an early multilayer optimizer run, evaluated at iterations 0, 15, and 30.

Three verdicts.

evaluator	endpoint ratio
independent FDTD	1.0007628
pre-repair analytical operator	3.6045899
pole-subtracted analytical operator	1.0000101

The direction was flat. The pre-repair operator reported a 3.6× improvement along it, which had it not been caught would have been an optimization campaign spent climbing an artefact. The repaired operator reproduces the flat verdict to five decimal places. Nothing was fitted or tuned to achieve this: the difference between the two analytical numbers is entirely the feed-pole treatment of §S10.

S18.7 Mesh-motion blockade

Approach. Apply the boundary velocity of (S69) to a meshed geometry with polygon repair, signed-distance reinitialization, and remeshing between steps.

Symptom. The line search stalled. The geometry actually delivered to the solver did not approach the previous geometry as the requested step went to zero, violating (S70), so no step size was small enough to produce a measurable, predictable change.

Resolution. The exact analytic parameterization of §S3, in which a step of size α moves the boundary by exactly α times the requested velocity. The mesh-based trajectory is retained as a counterexample.

S18.8 Short-ringdown quality factors

Approach. Fit Q from a 12 ps ringdown along a trajectory whose photon lifetime is growing.

Symptom. At step 61, $Q_{\mathrm{FDTD}} = 156\,520$ with $\mathcal{N} \approx 0.092$; the subsequent 64 ps analyses of geometries with strictly higher analytical Q return values near 58 240.27.

Resolution. Partial. The protocol was lengthened mid-campaign, which is why the two segments must be treated separately, and the adequacy audit of §S15 now quantifies the problem for every retained point. The full resolution is the matched panel, with the run-time amendment recorded there.

S18.9 Objectives that have not passed their own gates

Effective mode volume and coupling derivatives were implemented but have not passed the algebraic and local-surrogate tests of §S12 to the standard the Q derivative has. They are therefore excluded from the quasi-Newton objective entirely and appear in this work only as reported diagnostics. Quality factor, mode volume, and coupling do not automatically share a trust level, and treating them as if they did is a failure mode in its own right.

S19 Reproducibility and provenance

S19.1 How the numbers reach this document

No numeric value in the main text or in this document is typed into the LaTeX source. Three generators read the retained evidence and emit machine records; two exporters turn those records into LaTeX macros and tables; the manuscripts reference the macros.

generator	output	used by
`ringdown_adequacy.py`	`ringdown-adequacy.json`	§S15; main-text ringdown table
`geometry_metrics.py`	`geometry-metrics.json`	§S16, §S17
`surrogate_fidelity.py`	`surrogate-fidelity.json`	main-text fidelity table
`analytical_validation.py`	`analytical-validation.json`	§S14
`export_values.py`	`values.tex`	every quoted number
`export_tables.py`	`tables/*.tex`	every long table

Two consequences are worth stating. First, when a matched panel later replaces a provisional endpoint, the JSON changes, the exporters re-run, and the prose is untouched—there is exactly one substitution point. Second, `analytical_validation.py` does not merely copy values out of the historical gate records: it declares, for each value, the literal string that must appear in the cited file, and it fails rather than emitting anything if that string is absent. The manuscript therefore cannot drift away from the record it cites.

S19.2 Statistical rules declared in advance

Two rules were fixed before the corresponding audits were run and are stated in the source of the generators that apply them.

Fit-quality exclusion.
An FDTD observation is disqualified when its reported resonance error exceeds 0.01. Applied to the retained trajectory this excludes exactly one point of 33 (§S17). The rule refers only to fit quality and never to the resulting quality factor.

Adequacy threshold.
A ringdown fit spanning fewer than one amplitude lifetime is flagged. This is a flag, not an exclusion: all such points are retained, marked, and reported.

S19.3 Provenance of the theory sections

Sections S3–S12 are written self-contained and in one notation. Each is nonetheless a rewrite of a specific retained derivation phase, and the mapping is below. The final column is checked at build time.

SI	topic	retained derivation, under `derivations/`	present
S3	exact parameterized dielectric	`quan_loncar/fryett_q_series_phase1`	✓
S4	Lippmann–Schwinger, $\det_{\mathrm{reg}}$	`quan_loncar/fryett_q_series_phase1`	✓
S5	homogeneous reference Green tensor	`quan_loncar/fryett_q_series_phase2`	✓
S6	stratified reference Green tensor	`device23/multilayer_nanobeam_phase1`	✓
S7	Sommerfeld contour and quadrature	`quan_loncar/fryett_q_series_phase3`	✓
S8	Zernike–Legendre aperture basis	`quan_loncar/fryett_q_series_phase4`	✓
S9	symmetry reduction, array assembly	`quan_loncar/fryett_q_series_phase6`	✓
S10	feed-pole extraction and residues	`device23/multilayer_nanobeam_phase7`	✓
S11	continuation, pole search, mode identity	`quan_loncar/fryett_q_series_phase7`	✓
S12	implicit pole and Q derivatives	`qnm-boundary-gradient.md`	✓

Two notation hazards are worth repeating here, because a reader following the mapping into the raw record will meet them immediately. The air-clad lineage was developed against the encapsulated SiN benchmark of Fryett *et al.* [18]—$W = 450\,\mathrm{nm}$, $T = 330\,\mathrm{nm}$, $n_b = 2.0$, $n_c = 1.47$, $\lambda \approx 740\,\mathrm{nm}$—and not against the production air-clad silicon device (§S16). And several multilayer phases describe a superseded 25-hole, 75-coordinate device rather than the production 42-hole, 126-coordinate one (§S17). The mathematics is device-independent; the constants are not.

S19.4 Raw artifacts

Field arrays from the FDTD campaigns total roughly 11 GB and are retained outside version control, indexed by absolute source path, byte size, and SHA-256 checksum. The compact records used in this work—endpoint summaries, the trajectory series, the frozen geometry vectors, and the diagnostic metadata—are all in version control and are what every generator above reads. Reproducing the analysis in this document therefore requires no access to the raw arrays; reproducing the FDTD numbers themselves does, or requires re-running the simulations.

S19.5 Build

A single command audits the claims ledger and the provenance index, runs the test suite, regenerates every figure and table, compiles both documents with a pinned TEX distribution, and builds the documentation site. The tests pin the exact values quoted here, so a change to any record that is not accompanied by a change to the corresponding assertion fails the build. No build command is able to submit a paid simulation; the revalidation plan is stored with an explicit flag recording that submission is not permitted by the build, and the panel status is predeclared not submitted.

References

[1] Philippe Lalanne, Wei Yan, Kevin Vynck, Christophe Sauvan, and Jean-Paul Hugonin. Light interaction with photonic and plasmonic resonances. *Laser & Photonics Reviews*, 12(5):1700113, 2018.

[2] Philip Trost Kristensen, Kathrin Herrmann, Francesco Intravaia, and Kurt Busch. Modeling electromagnetic resonators using quasinormal modes. *Advances in Optics and Photonics*, 12(3):612–708, 2020.

[3] P. T. Leung, S. Y. Liu, and K. Young. Completeness and orthogonality of quasinormal modes in leaky optical cavities. *Physical Review A*, 49:3057–3067, 1994.

[4] Stefan Güttel and Françoise Tisseur. The nonlinear eigenvalue problem. *Acta Numerica*, 26:1–94, 2017.

[5] Alexander S. Markus. Introduction to the spectral theory of polynomial operator pencils. *Translations of Mathematical Monographs*, 71, 1988.

[6] Arthur D. Yaghjian. Electric dyadic Green's functions in the source region. *Proceedings of the IEEE*, 68(2):248–263, 1980.

[7] Barry Simon. *Trace Ideals and Their Applications.* American Mathematical Society, 2nd edition, 2005.

[8] Frank W. J. Olver, Daniel W. Lozier, Ronald F. Boisvert, and Charles W. Clark. *NIST Handbook of Mathematical Functions.* Cambridge University Press, 2010.

[9] Dennis P. Nyquist, D. R. Infante, and J. S. Bagby. Orthogonality and amplitude spectrum of radiation modes along open-boundary waveguides. *Journal of the Optical Society of America*, 71(1):49–54, 1981.

[10] Max Born and Emil Wolf. Principles of optics. 1999.

[11] Lloyd N. Trefethen and J. A. C. Weideman. The exponentially convergent trapezoidal rule. *SIAM Review*, 56(3):385–458, 2014.

[12] Steven G. Johnson, M. Ibanescu, M. A. Skorobogatiy, O. Weisberg, J. D. Joannopoulos, and Y. Fink. Perturbation theory for Maxwell's equations with shifting material boundaries. *Physical Review E*, 65:066611, 2002.

[13] Jan Sokołowski and Jean-Paul Zolésio. *Introduction to Shape Optimization: Shape Sensitivity Analysis*. Springer, 1992.

[14] Allen Taflove and Susan C. Hagness. *Computational Electrodynamics: The Finite-Difference Time-Domain Method*. Artech House, 3rd edition, 2005.

[15] Jean-Pierre Bérenger. A perfectly matched layer for the absorption of electromagnetic waves. *Journal of Computational Physics*, 114(2):185–200, 1994.

[16] Flexcompute. Tidy3D: GPU-accelerated finite-difference time-domain solver. https://www.flexcompute.com/tidy3d/, 2024.

[17] Qimin Quan and Marko Lončar. Deterministic design of wavelength scale, ultra-high Q photonic crystal nanobeam cavities. *Optics Express*, 19(19):18529–18542, 2011.

[18] Taylor K. Fryett, Yueyang Chen, James Whitehead, Zane M. Peycke, Xiaodong Xu, and Arka Majumdar. Encapsulated silicon nitride nanobeam cavity for hybrid nanophotonics. *ACS Photonics*, 5(6):2176–2181, 2018.