RESIDUAL-AWARE OUTPUT REPRESENTATIONS FOR SCIENTIFIC SURROGATES

Anonymous authors
Paper under double-blind review

ABSTRACT

Scientific surrogates are commonly trained to predict a complete physical response, whereas designs care about performance metrics derived from that response. This creates a distinct failure mode when a design-relevant metric is controlled by a small physical residual recovered through cancellation. We propose the residual-aware output representation method, where the surrogate exposes the small residual as an explicit output coordinate, reconstructs the complete response through an invertible transformation, and retains the original response-space training objective. In radio-frequency passives, exposing dissipation through impedance-like coordinates reduces relative Q error by $3.3\times$ on differential inductors and $6.7\times$ on six-port transformers with minimal change in scattering-response error; it also reduces circuit-level prediction violations. We further validate the proposed method on nanophotonic metasurfaces, reducing relative absorptance error. These results establish residual-aware output representation as a practical approach to cancellation-sensitive scientific problems.

1 INTRODUCTION

Scientific surrogates replace expensive numerical simulators by learning a map from design parameters x to a physical response y. They are commonly trained to minimize the error $\|\hat{y}(x) - y(x)\|$ over the complete response, so that the surrogate retains all physical information. Downstream designs, however, are selected through a design-relevant readout $g(\hat{y}(x))$ extracted from the surrogate response. The surrogate is therefore trained and used under different objectives: the response y during learning, and the readout metric $g(y)$ during design.

Small response error does not guarantee an accurate readout. The error magnitude $\delta = \hat{y} - y$ measures the average prediction error, but not how strongly it affects g, which depends to first order on $(\nabla g(y)^\top \delta)$. The gap is non-negligible when $g(y)$ is controlled by a small and design-critical residual $m(y)$ recovered through near-cancellation between larger components of y: an error negligible on the response scale can dominate the residual. We call this *directional error amplification*, because only the response-error component normal to the physical boundary is amplified to first order. The same gap between response accuracy and downstream accuracy arises in compound stability computed from formation-energy differences (Bartel et al., 2020), atmospheric heating rates from radiative-flux divergence (Ukkonen, 2022), weak radiated sound predicted from much larger flow fields (Tam, 1995), and aerodynamic drag estimated from predicted flow fields (Bonnet et al., 2022; Yagoubi et al., 2025).

RF passives provide an explicit geometric instance. Here, y is the scattering response $S(f)$ of an on-chip inductor or transformer, which takes the field solver minutes to hours to simulate. Design loops that need thousands to millions of evaluations (Settaluri et al., 2020; Wang et al., 2020) rely on surrogates that predict $S(f)$ in milliseconds (Fig. 1a). The performance metric g is the quality factor Q, the ratio of stored to dissipated energy, which can be mathematically derived from S. The response of a low-loss inductor lies near the unit circle of reflection S, and dissipation is represented in the narrow radial margin (Fig. 1b). While a tangential error in S changes only the reactance to first order, a radial error changes this margin and is amplified when Q is extracted, so the amplification is directional. A predicted S within 0.03 of the solver returns Q $14.8\times$ the ground truth (Fig. 1a) and mispredicts the performance of circuits built from it (Fig. 1c).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

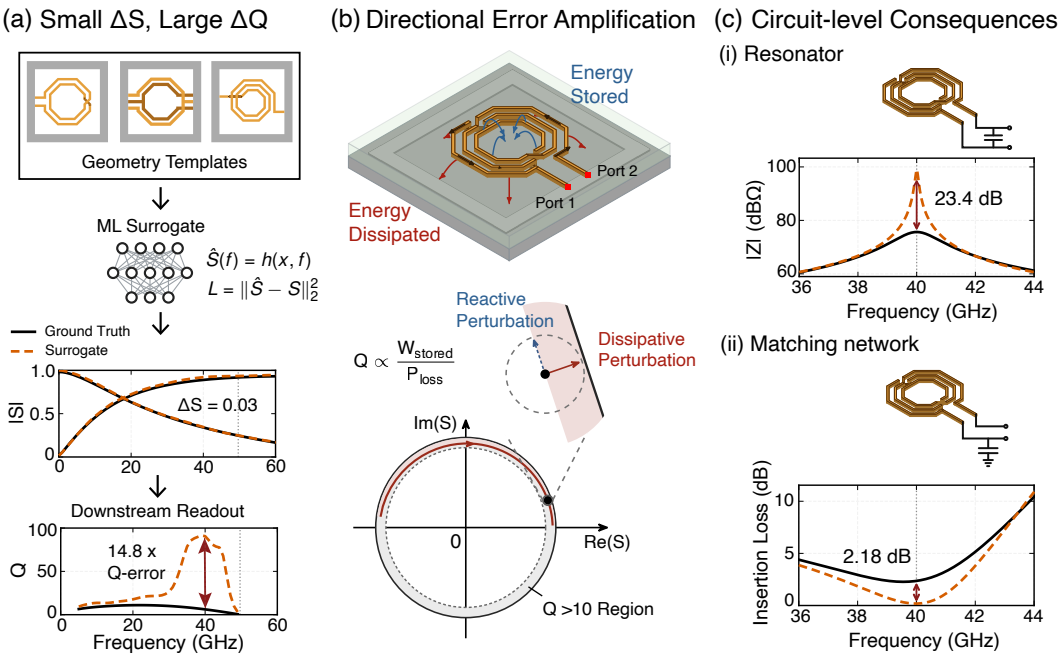

Figure 1: Small response error can conceal large downstream error. **(a)** For one differential inductor, an S prediction that differs from the solver by 0.03 gives a Q prediction $14.8\times$ the ground truth at 40 GHz. **(b)** Radial perturbations change the small dissipative margin $1 - |s_{\mathrm{d}}|^2$, whereas tangential perturbations do not to first order. **(c)** The same response error leads to large resonator-peak and matching-network insertion-loss discrepancies.

Motivated by the directional error amplification phenomenon, we propose the residual-aware output representation, by exposing the residual information directly to the output coordinates. We represent the complete response through an invertible map

$$z = \Phi(y) = \big(\rho(m(y)),\, u(y)\big),$$

where $m(y)$ is the small physical residual. The surrogate predicts $\hat{z}$, reconstructs $\hat{y} = \Phi^{-1}(\hat{z})$, and is trained by evaluating the original response-space loss on the reconstructed response. In RF passives, the exposed residual describes dissipation; in nanophotonic metasurfaces, it reflects the proportion of incident power absorbed rather than reflected or transmitted.

Our contributions are as follows.

1. We formulate the gap between response accuracy and task fidelity for cancellation-sensitive readout metrics, where a small residual of a complete physical response controls a design decision. We trace it to directional error amplification.

2. We introduce the residual-aware output representation method: the surrogate predicts an explicit residual coordinate and complementary coordinates, reconstructs the complete response through an invertible transformation, and is trained under the unchanged physical response loss. We show how this changes the local error geometry.

3. We distinguish representation-level intervention from other task-targeted methods, including task supervision, loss reweighting, and separate readout heads. Across RF passives and nanophotonic metasurfaces, residual-aware coordinates improve derived-metric prediction while retaining complete-response fidelity. At the circuit level, they predict realized performance more accurately and select designs closer to the optimum.

2 METHOD: RESIDUAL-AWARE OUTPUT REPRESENTATION

2.1 DIRECTIONAL ERROR AMPLIFICATION

Let $y \in \mathcal{Y} \subset \mathbb{R}^d$ denote a complete physical response, and let $m(y) > 0$ be a task-relevant physical residual that vanishes at a conservation, passivity, or feasibility boundary. We consider deployments whose accuracy is sensitive to relative variation in m.

Let $\delta = \hat{y} - y$ be a response perturbation. At a regular point with $\nabla m(y) \neq 0$, define the boundary-normal direction and the normal and tangential components of the perturbation by

$$\boldsymbol{n}_m(y) = \frac{\nabla m(y)}{\|\nabla m(y)\|_2}, \qquad \delta_{\mathrm{N}} = \boldsymbol{n}_m \boldsymbol{n}_m^\top \delta, \qquad \delta_{\mathrm{T}} = (I - \boldsymbol{n}_m \boldsymbol{n}_m^\top)\delta. \tag{1}$$

A first-order expansion gives

$$\Delta \log m = \frac{\|\nabla m(y)\|_2}{m(y)}\, \boldsymbol{n}_m(y)^\top \delta + o(\|\delta\|_2). \tag{2}$$

Equation (2) is directional: the tangential component δ_{T} does not change m to first order, whereas the normal component δ_{N} is amplified by $\|\nabla m\|_2/m$.

2.2 RESIDUAL-AWARE OUTPUT REPRESENTATION

Section 2.1 identifies the response-error direction amplified near the physical boundary. We therefore make the corresponding physical residual an explicit output coordinate.

Residual-exposing construction. Suppose a positive, task-critical residual $m(y)$ can be computed from the complete response and $u(y)$ contains the remaining information needed to reconstruct y from $m(y)$. Let $\rho(m) = \log m$. We construct

$$z = \Phi(y) = \big(\rho(m(y)),\, u(y)\big). \tag{3}$$

We choose u so that Φ is invertible on the relevant physical response domain. The log coordinate captures relative changes in the residual, since $\mathrm{d}\rho = \mathrm{d}m/m$. Any real predicted residual coordinate $\hat{\rho}$ yields $\hat{m} = e^{\hat{\rho}} > 0$ under the inverse map, enforcing residual positivity by construction. The surrogate predicts $\hat{z} = h_\theta(x)$, reconstructs $\hat{y} = \Phi^{-1}(\hat{z})$, and minimizes the original response-space loss $\mathcal{L}(\theta) = \mathbb{E}[\|\hat{y} - y\|_2^2]$.

Transformation geometry. Let $z^\star = \Phi(y)$ and $\delta z = \hat{z} - z^\star$, and let $J = J_{\Phi^{-1}}(z^\star)$ denote the Jacobian of the inverse transformation at $z^\star$. Locally, the inverse transformation maps coordinate learning error to physical response error as

$$\delta y \simeq J \delta z, \qquad \|\delta y\|_2^2 \simeq \delta z^\top J^\top J \delta z. \tag{4}$$

The corresponding readout error is

$$\Delta \log |g| \simeq \underbrace{\nabla_y \log |g(y)|^\top J}_{\text{readout sensitivity in residual-aware coordinates}} \delta z. \tag{5}$$

This expression evaluates the readout error along the learned coordinate-error direction δz. To compare representations independently of that learned direction, define the maximum directional readout amplification under a stated coordinate-error metric M_Φ, with $\|v\|_{M_\Phi}^2 = v^\top M_\Phi v$, as

$$\mathcal{A}_\Phi^{\max}(y) = \sup_{\|v\|_{M_\Phi}=1} \left|\nabla_y \log |g(y)|^\top J v\right| = \left\| M_\Phi^{-1/2} J^\top \nabla_y \log |g(y)| \right\|_2. \tag{6}$$

Making the physical residual an explicit coordinate separates boundary-normal variation from complementary response variation. For the parameterization $\rho = \log m$, $\mathrm{d}\rho = \mathrm{d}\log m$, so the residual coordinate records relative changes in m. The inverse Jacobian J maps errors in this coordinate and in u back to the physical response. The resulting output geometry therefore represents the cancellation-sensitive direction explicitly.

3 RF PASSIVES: DISSIPATION AND DIRECTIONAL ERROR AMPLIFICATION

Dissipative margin. RF field solvers return a complex scattering response $S(f)$ for a passive device. For a differential two-port structure, a fixed mode transformation produces the differential reflection coefficient $s_{\mathrm{d}} = a + jb$. Its dissipative margin $\varepsilon = 1 - |s_{\mathrm{d}}|^2$ is small in a low-loss device, while the quality factor is $Q = 2|b|/\varepsilon$; equations for full extraction are in Appendix A. For the componentwise-S baseline, the differential-mode prediction error is $\delta_s = \hat{s_{\mathrm{d}}} - s_{\mathrm{d}} = \delta_a + j\delta_b$. Because the level sets of ε are circles in the complex s_{d}-plane, the normal and tangential directions of Section 2.1 correspond here to radial and tangential directions.

We focus on the inductive region where $b > 0$, the intended operating points of the inductor and transformer templates studied here. Using $\Delta\varepsilon \simeq -2(a\delta_a + b\delta_b)$ and $Q = 2b/\varepsilon$, we derive

$$\begin{aligned} \Delta Q &\simeq \frac{2}{\varepsilon}\,\delta_b - \frac{2b}{\varepsilon^2}\,\Delta\varepsilon \\ &\simeq \frac{2}{\varepsilon}\,\delta_b + \frac{4b}{\varepsilon^2}(a\delta_a + b\delta_b) \\ &= \frac{Q}{b}\,\delta_b + \frac{Q^2}{b}(a\delta_a + b\delta_b). \end{aligned} \tag{7}$$

The direct-b term scales as Q. A tangential perturbation eliminates the second term but can retain the first, whereas a radial change in ε introduces the second term with a coefficient that scales as Q^2.

To diagnose the direction of the learned response error, we use

$$K \equiv |a\delta_a + b\delta_b|, \qquad T \equiv |a\delta_b - b\delta_a|.$$

The ratio K/T compares its radial and tangential content (Fig. 2b). For componentwise coordinates $z_S = (a, b)$ with $M_S = I$, Eq. 6 applied to $\log Q$ gives

$$\mathcal{A}^{\max}_{\Phi_S} > Q. \tag{8}$$

Thus, at fixed $K/|b|$, the relative error $|\Delta Q|/Q$ grows approximately linearly with Q: a high-Q (fabrication process dependent, typically > 10) response requires proportionally finer global accuracy (Fig. 4 and Appendix C).

Explicit dissipation. In impedance decomposition, R represents irreversible dissipation, and X represents reactive energy storage. For a scalar reflection mode s with positive reference impedance Z_0 ($100\,\Omega$ for the differential mode), we therefore use

$$\Phi(s) = (\log R, X), \qquad R + jX = Z_0\frac{1+s}{1-s}. \tag{9}$$

Here $R = Z_0\,\varepsilon/|1-s|^2$ vanishes exactly with the dissipative margin ε, so $\log R$ resolves relative loss variation while X retains the reactive degree of freedom. Because $\rho = \log R \in \mathbb{R}$ implies $R = e^{\rho} > 0$, every reconstructed scalar mode satisfies $1 - |s|^2 = 4RZ_0/|Z + Z_0|^2 > 0$. Thus each transformed reflection mode is contractive by construction (Proposition 1). We apply this transformation to the differential mode and retain all complementary scattering components, reconstructing the complete response; multiport mappings are given in Appendix A. For residual-aware coordinates $z_Z = (\rho, X)$ with $M_Z = \mathrm{diag}(1, X^{-2})$, Eq. 6 gives

$$\mathrm{d}\log|Q| = -\mathrm{d}\rho + \frac{\mathrm{d}X}{X}, \qquad \mathcal{A}^{\max}_{\Phi_Z} = \sqrt{2}. \tag{10}$$

Under the stated coordinate metrics, Q sensitivity is $O(1)$ in residual-aware coordinates but grows as $O(Q)$ in componentwise-S coordinates; Appendix C gives the exact componentwise expression.

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

Case Study I: RF passives. We evaluate 30,000 differential inductors with 2×2 responses, 30,000 transformers with 6×6 responses, and 30,000 single-ended spiral inductors with 2×2 responses.

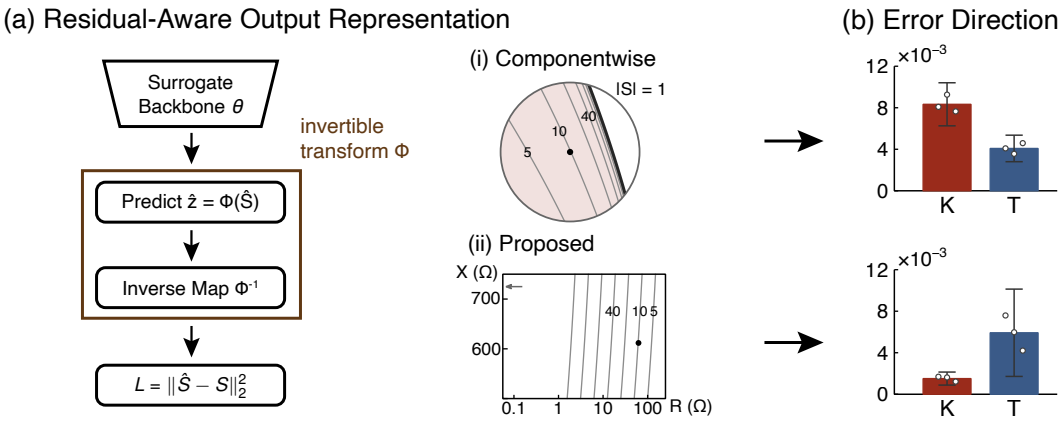

Figure 2: Residual-aware output representation in RF passives. **(a)** Residual-aware prediction pipeline: the network backbone predicts $\hat{z}$, and the fixed inverse map Φ^{-1} reconstructs $\hat{S}$ before the unchanged S-space loss is evaluated. Coordinate views (i) and (ii) show the same response in componentwise S and $(\log R, X)$, respectively: high-Q contours crowd near $|S| = 1$ in componentwise S, while the dissipative direction is explicit in $(\log R, X)$. **(b)** Radial and tangential prediction-error components on the differential-inductor test set. Averaged over three seed-level ratios, K/T is 2.08 for componentwise S and 0.26 for residual-aware coordinates.

Devices in each template are split $80/10/10\%$ into training, validation, and test sets and simulated at 201 frequencies. All RF surrogates use a latent-conditioned neural field over frequency (Xie et al., 2022): geometry is encoded into latent tokens, and a Fourier-feature frequency query (Mildenhall et al., 2020; Tancik et al., 2020) cross-attends to them (Jaegle et al., 2022) to emit the scattering response at that frequency. The process technology, electromagnetic solver, and open-PDK replication are documented in Appendices G and H.

Evaluation. We report scattering-response MAE, $\mathrm{MAE}(S)$, and relative Q error, relQ: for each device, we average $|\hat{Q} - Q|/|Q|$ over normalized frequency $u = f/f_{\mathrm{SRF}} \in [0.1, 1.0]$ and then take the median across devices; primary coil Q_{p} and secondary coil Q_{s} are reported separately for transformers. Thus, every reported Q is extracted from reconstructed $\hat{S}$; none of the representation comparisons receives Q supervision. Unless otherwise noted, we report the mean over three training seeds and 95% t-confidence intervals across seed-level metrics. Further details are in Appendices B and G, and more architecture information is in Appendix D.

4.2 Results across RF passive devices

On differential inductors, the proposed representation reduces relative Q error by $3.3\times$ with similar $\mathrm{MAE}(S)$ performance (Table 1). Fig. 2b resolves the associated change in error orientation: mean K decreases from 8.3×10^{-3} to 1.5×10^{-3}, while mean T remains at the same scale (4.1×10^{-3} and 5.9×10^{-3}); averaged over the three seed-level ratios, K/T shifts from 2.08 to 0.26. Over the evaluated band, the log–log Q error slope also changes from positive to negative.

The result extends beyond the differential two-port decomposition. The representation reduces relative Q_{p} and Q_{s} errors for six-port transformers by $6.7\times$ and $6.5\times$, respectively, and relative Q error for single-ended spiral inductors by $6.8\times$, while retaining comparable $\mathrm{MAE}(S)$ and reversing all error slopes (Table 1).

4.3 Alternatives to residual-aware representation

We compare the residual-aware output representation method with alternatives for targeting Q prediction: Q supervision, modified loss function, and a separate Q prediction head. All comparisons share the same dataset, network architecture and training schedule. The following experiments

Table 1: Effect of output representation across three device templates. The error slope is the slope of $\log(|\Delta Q|/Q)$ versus $\log Q$; a positive value indicates that relative error increases with Q.

output representation	readout	$\mathrm{MAE}(S)\downarrow$	relQ $\downarrow$	error slope
Differential Inductor, 2 *ports*				
componentwise S	Q	$\mathbf{0.00624}$ ± 0.00087	0.172 ± 0.014	$+0.41$
proposed	Q	0.00648 ± 0.00136	$\mathbf{0.053}$ ± 0.018	-0.34
Transformer, 6 *ports*				
componentwise S	Q_{p}	0.00629 ± 0.00039	0.725 ± 0.295	$+0.23$
	Q_{s}		0.738 ± 0.181	$+0.37$
proposed	Q_{p}	$\mathbf{0.00607}$ ± 0.00029	$\mathbf{0.109}$ ± 0.007	-0.45
	Q_{s}		$\mathbf{0.114}$ ± 0.004	-0.36
Single-ended Spiral Inductor, 2 *ports*				
componentwise S	Q	$\mathbf{0.02488}$ ± 0.00172	0.804 ± 0.141	$+0.83$
proposed	Q	0.02564 ± 0.00166	$\mathbf{0.119}$ ± 0.033	-0.23

Table 2: Auxiliary Q supervision with a componentwise-S output. The proposed representation is included for reference.

λ	$\mathrm{MAE}(S)\downarrow$	relQ $\downarrow$	$\mathrm{MAE}(S)$ vs λ=0
$\mathrm{MAE}(S)$ *at most* $2\times$ *the baseline*			
0	0.00624 ± 0.00087	0.172 ± 0.014	1.0
0.001	0.00636 ± 0.00185	0.113 ± 0.025	1.0
0.003	0.00598 ± 0.00208	0.117 ± 0.033	1.0
0.01	0.00637 ± 0.00075	0.098 ± 0.021	1.0
0.03	0.01007 ± 0.00210	0.125 ± 0.016	1.6
$\mathrm{MAE}(S)$ *above* $2\times$ *the baseline*			
0.1	0.05867 ± 0.00846	0.195 ± 0.042	9.4
0.3	0.06664 ± 0.00734	0.078 ± 0.030	10.7
1	0.12880 ± 0.00246	0.044 ± 0.036	20.6
proposed	0.00648 ± 0.00136	0.053 ± 0.018	1.0

show that the representation-level intervention achieves the highest Q accuracy derived from the complete-response prediction $\hat{S}$.

Objective level: Q supervision. We keep the componentwise-S output and add a $\log|Q|$ loss as an auxiliary task with weight λ (Caruana, 1997), where the predicted Q is extracted from the same $\hat{S}$ on which the response loss is computed. Because it is computed from $\hat{S}$, the auxiliary term directly penalizes the readout error amplified in Eq. 7. However, it leaves the componentwise output geometry unchanged, and dissipation is still represented implicitly. Table 2 shows the resulting trade-off: it reaches lower relative Q error than the proposed representation only after a large increase in $\mathrm{MAE}(S)$.

Metric level: impedance-aware response error. With componentwise-S outputs fixed, we weight each response error by the quadratic form $\delta_S^\top J_Z^\top J_Z \delta_S$, which approximates the induced squared error in normalized $(\log R, X)$ coordinates to first order and has the Jacobian-product structure of a Gauss–Newton metric (Botev et al., 2017; Martens, 2020). Across diagonal and full-matrix variants, the best relative Q error is 0.174, compared with 0.172 for the unweighted baseline. We also keep the $(\log R, X)$ output and evaluate the loss directly in impedance space, analogous to transformed-target regression (Box & Cox, 1964); the resulting $\mathrm{MAE}(S)$ values are 46–$50\times$ larger than with the reconstructed-S loss. Appendix E.2 gives the loss definitions and sweep results.

Prediction level: a separate Q head. A second head reaches relative Q error 0.0177, while Q extracted from the same models' predicted S remains between 0.155 and 0.253. Direct prediction therefore improves a separate supervised metric, not the response supplied to circuit composition (Appendix E.3).

Table 3: Circuit-level outcomes pooled across three training seeds. Reported values are medians except for percentages. Positive violation means that realized performance is worse than predicted; lower is better throughout. The filtered baseline restricts matching-network selection to candidates it predicts to be passive.

Design metric	componentwise S	componentwise S + filter	proposed representation
Transformer matching network			
insertion-loss violation (dB)	$+4.583$	$+0.767$	$\mathbf{+0.114}$
optimality gap (dB)	$+3.244$	$+0.277$	$\mathbf{+0.097}$
specification misses (%)	56.6	0.3	**0.0**
LC tank			
Leeson PN violation (dB)	$+2.35$	—	$\mathbf{+0.23}$
optimistic (%)	99.6	—	**66.8**

4.4 Downstream circuit design and selection

We test whether the response-level improvements influence circuit design tasks. We choose two circuits to optimize: sizing an LC tank with a differential inductor and selecting a transformer matching network from a finite library. Each surrogate first selects a circuit realization from its predicted response; we then keep that realization fixed and recompute its performance after replacing the predicted device response with the field-solver response of the same device. We define *prediction violation* as realized minus predicted performance, so a positive value means that the realized circuit is worse than predicted. For matching-network selection, we additionally report the rate at which the selected design misses the constraint that it was predicted to satisfy, and its *optimality gap* relative to the best solver-feasible candidate in the fully enumerated library. Appendix I gives the complete setup and discussion.

Oscillator tank. The phase noise of an LC oscillator depends on the quality factor of its tank (Leeson, 1966; Hajimiri & Lee, 1998). For each test differential inductor, the surrogate sizes an ideal capacitor from its predicted reactance at $f_{\mathrm{d}} = 0.5 f_{\mathrm{SRF}}$ and uses its predicted Q to calculate the phase noise. We then keep the capacitor and the remaining Leeson-model terms fixed, replace the predicted inductor response with its field-solver response, locate the resulting resonance, and recompute phase noise. As a result, the componentwise-S surrogate is optimistic for 99.6% of the tanks. The proposed representation reduces the median phase-noise violation from $+2.35$ to $+0.23$ dB, a $10.2\times$ reduction.

Transformer matching network. We use a differential-to-single-ended transformer with one shunt capacitor on each side, a standard topology at millimeter-wave frequencies (Vigilante & Reynaert, 2017; Er et al., 2021). We define insertion loss as $\mathrm{IL} = -20 \log_{10} |S_{21}|$. At 40 GHz and for each of $1{,}000$ source impedances sampled across the specified range, each surrogate searches the same transformer–capacitor library subject to a predicted $|S_{11}| < -10$ dB constraint and returns the candidate with the lowest predicted insertion loss. We evaluate that candidate under the solver response and compare it with the best solver-feasible candidate in the library.

Without filtering, the componentwise-S surrogate has a $40.2\times$ larger median insertion-loss violation and a $33.4\times$ larger optimality gap than the proposed representation. Its search frequently selects candidates that it predicts to be non-passive, so we also evaluate a deployable passivity filter. After filtering, the proposed representation still reduces violation by $6.7\times$ and optimality gap by $2.9\times$; specification misses fall from 0.3% to 0%. These results show that the residual-aware representation improves circuit-level prediction accuracy and effectiveness.

4.5 Case Study II: nanophotonic absorptance

To test whether the same representation principle transfers to a different physical domain, we consider the absorptance of nanophotonic metasurfaces. We use rigorous coupled-wave analysis (RCWA) (Moharam & Gaylord, 1981; Moharam et al., 1995) as the solver reference. For a fixed incident polarization, RCWA returns the power-normalized zeroth-order reflected and transmitted complex amplitudes $v = (r_x, r_y, \tilde{t}_x, \tilde{t}_y)$, and absorptance is the remainder after reflected and transmitted

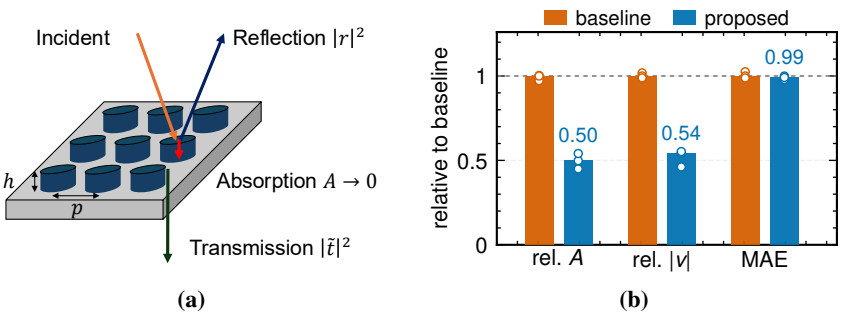

Figure 3: Case Study II: nanophotonic absorptance. **(a)** The metasurface geometry and partition of incident power: absorptance is the small residual left after reflection and transmission. **(b)** Errors relative to the componentwise-response baseline. Bars show three-seed medians, circles show individual seeds, and labels report proposed-to-baseline ratios.

power are accounted for:

$$A = 1 - |r|^2 - |\tilde{t}|^2 = 1 - \|v\|_2^2. \tag{11}$$

For a componentwise-response perturbation δv in the real embedding and $M_v = I$, Eq. 6 gives

$$\Delta \log A \simeq -\frac{2}{A}\,\mathrm{Re}(v^\dagger \delta v), \qquad \mathcal{A}^{\max}_{\Phi_v} = \frac{2\|v\|_2}{A} = \frac{2\sqrt{1-A}}{A}. \tag{12}$$

Thus, a perturbation tangent to the constant-$\|v\|_2$ surface leaves absorptance unchanged to first order, whereas a radial perturbation is amplified as $A \to 0$.

The componentwise baseline outputs the four complex amplitudes directly. The proposed representation instead uses the log-absorptance coordinate $\mu = -\log A$ and encodes the direction $v/\|v\|_2$ with a fixed invertible chart. Since $\Delta \log A = -\Delta\mu$, perturbations in the residual coordinate are not amplified as A approaches zero. The corresponding product metric $M_\mu = \mathrm{diag}(1, M_u)$ gives

$$\mathcal{A}^{\max}_{\Phi_\mu} = 1. \tag{13}$$

Both models reconstruct the same eight real-valued response components and are trained with the same loss definition in v-space. On the 48-wavelength grid, the proposed representation reduces median relative absorptance error by $2.01\times$, from 14.36% to 7.15%, while response MAE changes only from 6.1×10^{-4} to 6.0×10^{-4} (Fig. 3b). Thus, the same residual-aware construction improves the accuracy of a derived metric while maintaining the complete response accuracy.

5 Related Work

RF surrogates. Complete-response surrogates map device geometry to a broadband physical response and are trained in response space. For RF passives, this response is commonly represented by scattering parameters and learned using neural networks or rational and pole–residue models (Zhang & Gupta, 2000; Koziel & Leifsson, 2013; Gustavsen & Semlyen, 1999; Feng et al., 2016; Liu et al., 2005; Passos et al., 2023; Chae et al., 2024; He et al., 2026). When only a fixed readout is required, direct predictors instead regress downstream metrics such as inductance and quality factor (Mandal et al., 2008; Passos et al., 2017; Huang & Aghasi, 2026; Asano & Noda, 2018), while end-to-end synthesis maps specifications directly to geometry (Er et al., 2021; Lee et al., 2022; Wang et al., 2026). These task-specific formulations do not retain a broadband response that can be reused across circuit calculations.

Task-aware learning and evaluation. Task-aware methods connect predictions to downstream metrics either during training or evaluation. During training, value-aware loss functions, predict-then-optimize, and decision-focused learning use downstream value or decision quality (Farahmand et al., 2017; Elmachtoub & Grigas, 2022; Mandi et al., 2024). TaskMet learns a prediction-space metric from a downstream task (Bansal et al., 2023); auxiliary supervision and multi-task loss weighting add derived targets to the response objective (Caruana, 1997; Kendall et al., 2018; Chen et al., 2018);

and transformed-target regression, including Box–Cox regression, evaluates the fit after changing the response scale (Box & Cox, 1964). During evaluation, metrics derived from predicted responses complement aggregate response error. Prior work reports RF specification thresholds (He et al., 2026), compound stability (Bartel et al., 2020), atmospheric heating rates (Ukkonen, 2022), and aerodynamic force or ranking metrics (Bonnet et al., 2022; Yagoubi et al., 2025). Recent discussions of scientific-model predictivity likewise distinguish response accuracy from accuracy in decisive metrics such as dissipation or phase (Duraisamy, 2026). The same metric-centered perspective underlies goal-oriented error estimation, which weights discretization error by its effect on a specified output functional (Becker & Rannacher, 2001; Venditti & Darmofal, 2000).

Structured output representations. Output coordinates can enforce physical structure, avoid discontinuities, or encode continuous targets differently. Deep Lagrangian Networks construct positive-definite mass matrices from lower-triangular factors (Lutter et al., 2019); continuous rotation representations avoid low-dimensional discontinuities (Zhou et al., 2019); and binary-encoded labels use multi-bit codes for continuous targets (Shah et al., 2022). In RF modeling, equivalent-circuit and transfer-function-coefficient models predict a compact parameterization of a prescribed passive template and reconstruct it algebraically (Cao & Wang, 2007; Zhao et al., 2020). These parameterizations restrict or compress the response and are not generally invertible on arbitrary broadband responses.

6 Limitations

The experiments cover two physical domains, while extending the method to substantially higher-dimensional response spaces remains an important direction. The method applies when a complete response contains a low-dimensional residual that controls the downstream metric. In RF, the current mapping guarantees contractivity for the represented reflection modes; extending this guarantee to complete multiport passivity, causality, and stability remains future work.

7 Conclusion

Output representation shapes which response-error directions a surrogate learns most accurately. For RF passives, making dissipation explicit shifts the learned error away from the Q-sensitive radial direction, improving quality-factor prediction while retaining scattering-response accuracy. Surrogates utilizing this representation support more accurate circuit prediction and design selection. The same construction improves absorptance prediction in metasurfaces, where absorptance is the residual of reflected and transmitted power.

These results suggest a practical design principle for complete-response surrogates: when a downstream metric is governed by cancellation, expose the associated physical residual in the output coordinates and evaluate the loss after reconstructing the original response.

AI use statement

Large language models, including Claude Code and Codex, were used to polish the language and assist with coding. The authors determined the scientific content, performed and verified all experiments, and checked the numerical results manually. The authors reviewed and take responsibility for the complete manuscript.

Reproducibility statement

Section 4.1 and Appendices B and G specify the training and evaluation setups; Appendices I and J describe the circuit-level and metasurface experiments. The commercial solver data and foundry PDK behind the main RF results cannot be redistributed. We replicate the single-ended spiral inductor template using an open PDK and an open field solver in Appendix H. The layout-generation and training code for this experiment will be shared with the reviewers and area chairs through an anonymous repository posted in the discussion forum on OpenReview.

REFERENCES

Amazon Web Services. Palace: 3D finite element solver for computational electromagnetics. `https://github.com/awslabs/palace`, 2026. Version 0.16.0, Apache License 2.0.

Takashi Asano and Susumu Noda. Optimization of photonic crystal nanocavities based on deep learning. *Optics Express*, 26(25):32704–32717, 2018. doi: 10.1364/OE.26.032704.

Dishank Bansal, Ricky T. Q. Chen, Mustafa Mukadam, and Brandon Amos. TaskMet: Task-driven metric learning for model learning. In *Advances in Neural Information Processing Systems 36 (NeurIPS 2023)*, 2023. doi: 10.52202/075280-2017.

Christopher J. Bartel, Amalie Trewartha, Qi Wang, Alexander Dunn, Anubhav Jain, and Gerbrand Ceder. A critical examination of compound stability predictions from machine-learned formation energies. *npj Computational Materials*, 6:97, 2020. doi: 10.1038/s41524-020-00362-y.

Roland Becker and Rolf Rannacher. An optimal control approach to a posteriori error estimation in finite element methods. *Acta Numerica*, 10:1–102, 2001. doi: 10.1017/S0962492901000010.

Florent Bonnet, Jocelyn Mazari, Paola Cinnella, and Patrick Gallinari. AirfRANS: High fidelity computational fluid dynamics dataset for approximating Reynolds-averaged Navier–Stokes solutions. In *Advances in Neural Information Processing Systems 35: Datasets and Benchmarks Track*, 2022. doi: 10.52202/068431-1705.

Aleksandar Botev, Hippolyt Ritter, and David Barber. Practical Gauss-Newton optimisation for deep learning. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, volume 70 of *Proceedings of Machine Learning Research*, pp. 557–565, 2017.

G. E. P. Box and D. R. Cox. An analysis of transformations. *Journal of the Royal Statistical Society: Series B*, 26(2):211–243, 1964. doi: 10.1111/j.2517-6161.1964.tb00553.x.

Yazi Cao and Gaofeng Wang. A wideband and scalable model of spiral inductors using space-mapping neural network. *IEEE Transactions on Microwave Theory and Techniques*, 55(12):2473–2480, 2007. doi: 10.1109/TMTT.2007.909602.

Rich Caruana. Multitask learning. *Machine Learning*, 28:41–75, 1997. doi: 10.1023/A:1007379606734.

Hyunsu Chae, Hao Yu, Sensen Li, and David Z. Pan. PulseRF: Physics-augmented ML modeling and synthesis for high-frequency RFIC design. In *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2024. doi: 10.1145/3676536.3676768.

Zhao Chen, Vijay Badrinarayanan, Chen-Yu Lee, and Andrew Rabinovich. GradNorm: Gradient normalization for adaptive loss balancing in deep multitask networks. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*, 2018.

Karthik Duraisamy. Predictivity and utility of neural surrogates of multiscale PDEs. *arXiv preprint arXiv:2604.20061*, 2026.

Adam N. Elmachtoub and Paul Grigas. Smart "predict, then optimize". *Management Science*, 68(1):9–26, 2022. doi: 10.1287/mnsc.2020.3922.

Siawpeng Er, Edward Liu, Minshuo Chen, Yan Li, Yuqi Liu, Tuo Zhao, and Hua Wang. Deep learning assisted end-to-end synthesis of mm-wave passive networks with 3d em structures: A study on a transformer-based matching network. In *IEEE MTT-S International Microwave Symposium (IMS)*, pp. 66–69, 2021.

Amir-massoud Farahmand, André M. S. Barreto, and Daniel N. Nikovski. Value-aware loss function for model-based reinforcement learning. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 54 of *Proceedings of Machine Learning Research*, pp. 1486–1494, 2017.

Feng Feng, Chao Zhang, Jianguo Ma, and Qi-Jun Zhang. Parametric modeling of EM behavior of microwave components using combined neural networks and pole-residue-based transfer functions. *IEEE Transactions on Microwave Theory and Techniques*, 64(1):60–77, 2016. doi: 10.1109/TMTT.2015.2504099.

B. Gustavsen and A. Semlyen. Rational approximation of frequency domain responses by vector fitting. *IEEE Transactions on Power Delivery*, 14(3):1052–1061, 1999. doi: 10.1109/61.772353.

Ali Hajimiri and Thomas H. Lee. A general theory of phase noise in electrical oscillators. *IEEE Journal of Solid-State Circuits*, 33(2):179–194, 1998.

Houbo He, Yizhou Xu, Lei Xia, Yaolong Hu, Fan Cai, and Taiyun Chi. MOTIF-RF: Multi-template on-chip transformer synthesis incorporating frequency-domain self-transfer learning for RFIC design automation. In *Asia and South Pacific Design Automation Conference (ASP-DAC)*, 2026. doi: 10.1109/ASP-DAC66049.2026.11420464.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016. doi: 10.1109/CVPR.2016.90.

Yilun Huang and Hamidreza Aghasi. AI-assisted radio-frequency and mm-wave layout using electromagnetic surrogates and frequency-adaptive rules. *IEEE Journal of Selected Topics in Electromagnetics, Antennas and Propagation*, 2:445–461, 2026. doi: 10.1109/JSTEAP.2026.3722758.

IHP GmbH. IHP open source PDK: SG13G2. `https://github.com/IHP-GmbH/IHP-Open-PDK`, 2026. Apache License 2.0; commit 22f2a25.

Andrew Jaegle, Sebastian Borgeaud, Jean-Baptiste Alayrac, Carl Doersch, Catalin Ionescu, David Ding, Skanda Koppula, Daniel Zoran, Andrew Brock, Evan Shelhamer, Olivier Hénaff, Matthew M. Botvinick, Andrew Zisserman, Oriol Vinyals, and João Carreira. Perceiver IO: A general architecture for structured inputs and outputs. In *International Conference on Learning Representations (ICLR)*, 2022.

Alex Kendall, Yarin Gal, and Roberto Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7482–7491, 2018. doi: 10.1109/CVPR.2018.00781.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015.

Slawomir Koziel and Leifur Leifsson (eds.). *Surrogate-Based Modeling and Optimization: Applications in Engineering*. Springer, New York, 2013. doi: 10.1007/978-1-4614-7551-4.

D. Lee, G. Shin, S. Lee, K. Kim, T.-H. Oh, and H.-J. Song. Neural-network-based automated synthesis of transformer matching circuits for rf amplifier design. *IEEE Transactions on Microwave Theory and Techniques*, 70(11):4726–4739, 2022. doi: 10.1109/TMTT.2022.3199756.

D. B. Leeson. A simple model of feedback oscillator noise spectrum. *Proceedings of the IEEE*, 54 (2):329–330, 1966.

Tao Liu, Wenjun Zhang, and Zhiping Yu. Modeling of spiral inductors using artificial neural network. In *2005 IEEE International Joint Conference on Neural Networks (IJCNN)*, volume 4, pp. 2353–2358, 2005. doi: 10.1109/IJCNN.2005.1556269.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations (ICLR)*, 2019.

Michael Lutter, Christian Ritter, and Jan Peters. Deep lagrangian networks: Using physics as model prior for deep learning. In *International Conference on Learning Representations (ICLR)*, 2019.

S. K. Mandal, S. Sural, and A. Patra. ANN- and PSO-based synthesis of on-chip spiral inductors for RF ICs. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 27(1): 188–192, 2008. doi: 10.1109/TCAD.2007.907284.

Jayanta Mandi, James Kotary, Senne Berden, Maxime Mulamba, Victor Bucarey, Tias Guns, and Ferdinando Fioretto. Decision-focused learning: Foundations, state of the art, benchmark and future opportunities. *Journal of Artificial Intelligence Research*, 80:1623–1701, 2024. doi: 10.1613/jair.1.15320.

James Martens. New insights and perspectives on the natural gradient method. *Journal of Machine Learning Research*, 21(146):1–76, 2020.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. In *European Conference on Computer Vision (ECCV)*, 2020. doi: 10.1007/978-3-030-58452-8_24.

M. G. Moharam and T. K. Gaylord. Rigorous coupled-wave analysis of planar-grating diffraction. *Journal of the Optical Society of America*, 71(7):811–818, 1981. doi: 10.1364/JOSA.71.000811.

M. G. Moharam, T. K. Gaylord, Eric B. Grann, and Drew A. Pommet. Formulation for stable and efficient implementation of the rigorous coupled-wave analysis of binary gratings. *Journal of the Optical Society of America A*, 12(5):1068–1076, 1995. doi: 10.1364/JOSAA.12.001068.

F. Passos, E. Roca, R. Castro-López, and F. V. Fernández. Radio-frequency inductor synthesis using evolutionary computation and Gaussian-process surrogate modeling. *Applied Soft Computing*, 60: 495–507, 2017. doi: 10.1016/j.asoc.2017.07.036.

Fábio Passos, Nuno Lourenço, Elisenda Roca, Ricardo Martins, Rafael Castro-López, Nuno Horta, and Francisco V. Fernández. PACOSYT: A passive component synthesis tool based on machine learning and tailored modeling strategies towards optimal RF and mm-wave circuit designs. *IEEE Journal of Microwaves*, 3(2):599–613, 2023. doi: 10.1109/JMW.2023.3237260.

Keertana Settaluri, Ameer Haj-Ali, Qijing Huang, Kourosh Hakhamaneshi, and Borivoje Nikolic. AutoCkt: Deep reinforcement learning of analog circuit designs. In *Design, Automation and Test in Europe Conference (DATE)*, pp. 490–495, 2020. doi: 10.23919/DATE48585.2020.9116200.

Deval Shah, Zi Yu Xue, and Tor M. Aamodt. Label encoding for regression networks. In *International Conference on Learning Representations (ICLR)*, 2022.

Christopher K. W. Tam. Computational aeroacoustics: Issues and methods. *AIAA Journal*, 33(10): 1788–1796, 1995. doi: 10.2514/3.12728.

Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan T. Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. In *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*, 2020.

Peter Ukkonen. Exploring pathways to more accurate machine learning emulation of atmospheric radiative transfer. *Journal of Advances in Modeling Earth Systems*, 14:e2021MS002875, 2022. doi: 10.1029/2021MS002875.

David A. Venditti and David L. Darmofal. Adjoint error estimation and grid adaptation for functional outputs: Application to quasi-one-dimensional flow. *Journal of Computational Physics*, 164(1): 204–227, 2000. doi: 10.1006/jcph.2000.6600.

Marco Vigilante and Patrick Reynaert. On the design of wideband transformer-based fourth order matching networks for e-band receivers in 28-nm cmos. *IEEE Journal of Solid-State Circuits*, 52 (8):2071–2082, 2017.

Hanrui Wang, Kuan Wang, Jiacheng Yang, Linxiao Shen, Nan Sun, Hae-Seung Lee, and Song Han. GCN-RL circuit designer: Transferable transistor sizing with graph neural networks and reinforcement learning. In *2020 57th ACM/IEEE Design Automation Conference (DAC)*, pp. 1–6, 2020. doi: 10.1109/DAC18072.2020.9218757.

Zuojun Wang, Jingyun Bi, Qin Dong, Zhou Shu, Jixin Chen, Wei Hong, Kevin Tshun Chuan Chai, and Yongxin Guo. Deep-neural-network-based automated transformer design for multistage wideband millimeter-wave power amplifiers. *IEEE Transactions on Microwave Theory and Techniques*, 2026. Early access.

Yiheng Xie, Towaki Takikawa, Shunsuke Saito, Or Litany, Shiqin Yan, Numair Khan, Federico Tombari, James Tompkin, Vincent Sitzmann, and Srinath Sridhar. Neural fields in visual computing and beyond. *Computer Graphics Forum*, 41(2):641–676, 2022. Eurographics State of the Art Report.

Mouadh Yagoubi, David Danan, Milad Leyli-Abadi, Ahmed Mazari, Jean-Patrick Brunet, Abbas Kabalan, Fabien Casenave, Yuxin Ma, Giovanni Catalani, Jean Fesquet, Jacob Helwig, Xuan Zhang, Haiyang Yu, Xavier Bertrand, Frederic Tost, Michael Bauerheim, Joseph Morlier, and Shuiwang Ji. NeurIPS 2024 ML4CFD competition: Results and retrospective analysis. *arXiv preprint arXiv:2506.08516*, 2025.

Qi-Jun Zhang and Kuldip C. Gupta. *Neural Networks for RF and Microwave Design*. Artech House, Norwood, MA, 2000.

Zhihao Zhao, Feng Feng, Wei Zhang, Jianan Zhang, Jing Jin, and Qi-Jun Zhang. Parametric modeling of EM behavior of microwave components using combined neural networks and hybrid-based transfer functions. *IEEE Access*, 8:93922–93938, 2020. doi: 10.1109/ACCESS.2020.2990157.

Yi Zhou, Connelly Barnes, Jingwan Lu, Jimei Yang, and Hao Li. On the continuity of rotation representations in neural networks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

A NOTATION, MODAL CONVENTIONS, AND EXTRACTION FORMULAS

Differential inductors. The reciprocal devices considered here return a 2×2 complex scattering matrix $S(f)$ referenced to $50\,\Omega$ per port. With the differential-mode reflection coefficient $s_{\mathrm{d}} = \frac{1}{2}(S_{11} + S_{22} - 2S_{12})$ of §3 referenced to $Z_0 = 100\,\Omega$, $s_{\mathrm{d}} = a + jb$ and $\varepsilon = 1 - |s_{\mathrm{d}}|^2 = 1 - a^2 - b^2$, multiplying numerator and denominator of $Z = Z_0(1 + s_{\mathrm{d}})/(1 - s_{\mathrm{d}})$ by $\overline{(1 - s_{\mathrm{d}})}$ gives

$$Z = Z_0 \, \frac{(1 - |s_{\mathrm{d}}|^2) + 2jb}{|1 - s_{\mathrm{d}}|^2} = Z_0 \, \frac{\varepsilon}{|1 - s_{\mathrm{d}}|^2} + j \, Z_0 \, \frac{2b}{|1 - s_{\mathrm{d}}|^2}, \tag{14}$$

hence $R = Z_0\varepsilon/|1 - s_{\mathrm{d}}|^2$, $X = 2Z_0 b/|1 - s_{\mathrm{d}}|^2$, $L = X/\omega$, and $Q = |X|/R = 2|b|/\varepsilon$. The common mode s_{c} is referenced to $25\,\Omega$ and is not used for the lumped differential quantities considered here. Passivity implies $\varepsilon \geq 0$ and therefore $R \geq 0$. The sensitivity analyzed in §3 arises because this nonnegative dissipation is encoded by the small margin ε.

Six-port transformer. The physical port order is $(p_+, p_-, p_{\mathrm{CT}}, s_+, s_-, s_{\mathrm{CT}})$, with each port referenced to $50\,\Omega$. For incident and reflected port-wave vectors, we apply the same orthogonal matrix

$$M = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 & 0 & 0 \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 0 & 0 & 0 & 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \qquad S_{\mathrm{m}} = MSM^{\mathsf{T}},$$

whose modal order is $(d_{\mathrm{p}}, c_{\mathrm{p}}, t_{\mathrm{p}}, d_{\mathrm{s}}, c_{\mathrm{s}}, t_{\mathrm{s}})$; the corresponding reference impedances are $(100, 25, 50, 100, 25, 50)\,\Omega$. Let S_{ps} be the 2×2 submatrix of S_{m} on modes $(d_{\mathrm{p}}, d_{\mathrm{s}})$. We define

$$100\,\Omega\,(I + S_{\mathrm{ps}})(I - S_{\mathrm{ps}})^{-1} = \begin{bmatrix} R_{\mathrm{p}} + jX_{\mathrm{p}} & Z_{\mathrm{ps}} \\ Z_{\mathrm{ps}} & R_{\mathrm{s}} + jX_{\mathrm{s}} \end{bmatrix}.$$

The six real coordinates of this block are $(\log R_{\mathrm{p}}, X_{\mathrm{p}}, \mathrm{Re}\, Z_{\mathrm{ps}}, \mathrm{Im}\, Z_{\mathrm{ps}}, \log R_{\mathrm{s}}, X_{\mathrm{s}})$. All other reciprocity-unique entries of S_{m}, including differential–common/tap couplings, remain in real–imaginary coordinates. The inverse map converts the impedance block back to S_{ps}, inserts it into the complete modal matrix, and applies $S = M^{\mathsf{T}} S_{\mathrm{m}} M$; no response entries are discarded. Finally, $L_{\mathrm{p,s}} = X_{\mathrm{p,s}}/\omega$, $Q_{\mathrm{p,s}} = |X_{\mathrm{p,s}}|/R_{\mathrm{p,s}}$, and the coupling coefficient is $k_{\mathrm{cpl}} = \mathrm{Im}(Z_{\mathrm{ps}})/\sqrt{|X_{\mathrm{p}} X_{\mathrm{s}}|}$. The transform and extraction convention are held fixed across representations, seeds, and transformer experiments.

Single-ended spiral inductors. These reciprocal two-ports are referenced to $50\,\Omega$ per port. We short port 2 ($\Gamma_2 = -1$) and apply Eq. 9, with $Z_0 = 50\,\Omega$, to the input reflection

$$s_{\mathrm{in}} = S_{11} - \frac{S_{12} S_{21}}{1 + S_{22}}.$$

The remaining independent entries $S_{12} = S_{21}$ and S_{22} are retained in real–imaginary coordinates, and the inverse map recovers $S_{11} = s_{\mathrm{in}} + S_{12}^2/(1 + S_{22})$. Thus the complete reciprocal two-port is reconstructed without a mode transform.

B EVALUATION PROTOCOL

In-band. For the differential-inductor results, the evaluation grid has 201 frequencies from 0 to 200 GHz in 1 GHz steps. If the true reactance first changes from positive to nonpositive between adjacent grid points f_k and f_{k+1}, we use $f_{\mathrm{SRF}} = f_k$, the last positive-reactance grid point, as the operational in-band cutoff; we do not interpolate the zero crossing. If no such change occurs on the grid, we use $f_{\mathrm{SRF}} = 200\,\mathrm{GHz}$. The normalized evaluation frequency is $u = f/f_{\mathrm{SRF}}$. For devices with a detected sign change, the included point at $u = 1$ precedes that change rather than being a point where the true Q vanishes.

The primary relative Q error metric is defined as

$$\mathrm{relQ} = \mathrm{median}_{\mathrm{devices}} \ \mathrm{mean}_{u\in[0.1,\,1.0]} \frac{|\hat{Q}(f) - Q(f)|}{|Q(f)|}, \tag{15}$$

where the mean is over original grid points and DC is excluded. The implementation uses $\max(|Q|, 10^{-12})$ as a numerical denominator guard and excludes points with $Q = 0$. Scattering-response MAE is $\mathrm{MAE}(S) = \mathrm{mean}\,|\hat{y} - y|$ over devices, frequencies, independent scattering entries, and real/imaginary parts in physical scattering units; each reciprocal scattering pair is counted once. The train-validation-test split is device-wise: all 201 frequencies of a device remain in the same split.

The absolute Q errors are also recorded over seeds, with an average of 1.521 for componentwise S and 0.354 for the proposed representation. Direct (R, X) output, which exposes the same dissipation coordinate without the logarithm, gives relative Q error 0.0576 and $\mathrm{MAE}(S) = 0.00651$ under the same training setup, compared with 0.0525 and 0.00648 for $(\log R, X)$. The accuracy gain therefore comes from exposing R as an output coordinate, while the logarithm additionally guarantees $R > 0$ and hence the contractivity of Proposition 1.

C First-order error analysis and mechanism diagnostic

Exact componentwise sensitivity. On the inductive branch $b > 0$, differentiating $\log Q$ gives

$$\mathrm{d}\log Q = \frac{Qa}{b}\,\mathrm{d}a + \left(Q + \frac{1}{b}\right)\mathrm{d}b.$$

For the stated componentwise coordinate metric $M_S = I$, the exact maximum directional amplification is therefore

$$\mathcal{A}^{\max}_{\Phi_S} = \sqrt{\left(\frac{Qa}{b}\right)^2 + \left(Q + \frac{1}{b}\right)^2} > Q. \tag{16}$$

Margin-term dominance check. We compare the absolute magnitudes of the margin-induced Q^2 term and the direct-b Q term in Eq. 7 using

$$\eta_Q \equiv \frac{\left|\frac{Q^2}{b}(a\delta_a + b\delta_b)\right|}{\left|\frac{Q}{b}\delta_b\right|}.$$

Over the in-band points, the margin-induced term exceeds the direct-b term at 99.3% of points, with median ratio $\eta_Q = 17.0$. Thus, the Q^2 contribution dominates the baseline's first-order Q error at most test points.

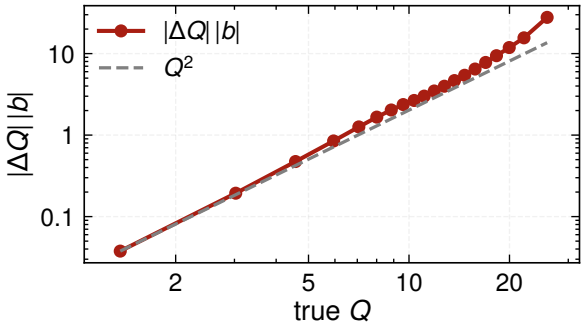

Figure 4: Mechanism diagnostic consistent with Eq. 7. Across in-band test points, the componentwise-S surrogate's $|\Delta Q|\,|b|$ follows the quadratic reference guide.

Table 4: Comparison of parameter-matched network backbones using $\mathrm{MAE}(S)$ and relQ in componentwise-S coordinates.

network backbone	parameters	$\mathrm{MAE}(S)\downarrow$	$\mathrm{relQ}\downarrow$
frequency-conditioned network backbone	19.36 M	**0.00627**	0.18663
flat residual network	19.13 M	0.01499	**0.14372**

D Architecture comparison

Frequency-conditioned network backbone. A common RF surrogate emits the full spectrum at once: a flat network maps the geometry vector to real and imaginary components at each of the 201 frequencies and for every independent scattering entry (Zhang & Gupta, 2000; Koziel & Leifsson, 2013; Feng et al., 2016). Such a model has no parameter sharing across frequency, and its output width is tied to the evaluation grid. Our network backbone instead uses a shared output map at queried frequencies.

Architecture comparison. Table 4 compares the frequency-conditioned network backbone with a parameter-matched flat residual network (He et al., 2016) in native componentwise-S coordinates. For this comparison, the frequency-conditioned backbone is widened to match the flat network's parameter count; the main representation experiments use its standard width. The frequency-conditioned network backbone has lower $\mathrm{MAE}(S)$, whereas the flat network has lower relative Q error derived from the response. Thus, selection by response error and selection by the downstream readout give different architecture rankings.

E Loss and prediction controls

All ablations in this section use test differential inductors and the common three-seed training protocol unless explicitly marked as a single run.

E.1 Auxiliary Q supervision

The objective-level control retains the componentwise-S output and adds a loss on Q extracted from that same response,

$$\mathcal{L}_\lambda = \mathcal{L}_S + \lambda \ \mathrm{mean}\left|\log|\hat{Q}(\hat{S})| - \log|Q(S)|\right|.$$

This control adds no output head: $\hat{S}$ remains the emitted response.

E.2 Impedance-aware response errors

These impedance-aware ablations try to get the benefit of the proposed coordinates through the loss rather than through the output representation. The Jacobian uses the dimensionless form $z_Z = (\log(R/Z_0), X/Z_0)$, which is the $(\log R, X)$ of the main text with the origin of $\log R$ shifted. Let $\delta_S = \mathrm{vec}_{\mathbb{R}}(\hat{S} - S)$ and let $J_Z = \partial z_Z/\partial\,\mathrm{vec}_{\mathbb{R}}(S)$ be the real Jacobian of the differential-mode impedance extraction. The local relation

$$\left\|\Delta z_Z\right\|_2^2 \simeq \delta_S^\top J_Z^\top J_Z\,\delta_S$$

connects the impedance-space loss to the full Jacobian weighting to first order.

Jacobian-weighted response loss. With the componentwise-S output fixed, let $G = J_Z^\top J_Z$, which has the form of the Gauss–Newton matrix (Botev et al., 2017; Martens, 2020) with the Jacobian taken from S to the impedance coordinates rather than from the parameters to the network outputs, and normalize it as $\widetilde{G} = G/\mathbb{E}_{\mathcal{B}}[\mathrm{tr}(G)]$ over each training minibatch $\mathcal{B}$. We add $\beta\,\mathbb{E}[\delta_S^\top \widetilde{G}\delta_S]$ to the unweighted response MSE. The diagonal form replaces G by $\mathrm{diag}(G)$, while the full form retains its cross terms. Across both forms with $\beta \in \{1, 10\}$, the relative Q error ranges from 0.174 to 0.194, compared with 0.172 for the unweighted baseline under the same training setup.

Table 5: A second $\log|Q|$ head attached to the componentwise-S surrogate. relQ_S is extracted from the predicted response and relQ_H is read from the additional head.

λ	$\mathrm{MAE}(S)\downarrow$	$\mathrm{relQ}_S\downarrow$	$\mathrm{relQ}_H\downarrow$
0.003	$0.00605\ \pm 0.00053$	$0.169\ \pm 0.086$	$0.0252\ \pm 0.0049$
0.01	$0.00651\ \pm 0.00159$	$0.155\ \pm 0.045$	$0.0204\ \pm 0.0082$
0.03	$0.00691\ \pm 0.00188$	$0.167\ \pm 0.048$	$0.0177\ \pm 0.0003$
0.3	$0.00929\ \pm 0.00070$	$0.253\ \pm 0.105$	$0.0221\ \pm 0.0105$

Scalar sensitivity weighting. We also retain only the scalar gain $w = |\partial Z/\partial s_{\mathrm{d}}|^2 = \left(2Z_0/|1 - s_{\mathrm{d}}|^2\right)^2$ as a pointwise weight on the componentwise-S residual. The weight is evaluated at the target, detached from the gradient, normalized to unit mean, and clamped at a multiple of the batch median. Across clipping thresholds 10^2, 10^3, and 10^4 times the batch median, scalar weighting increases both $\mathrm{MAE}(S)$ and relQ relative to the unweighted baseline.

Loss evaluated in impedance space. We keep the $(\log R, X)$ output representation and replace the reconstructed-S loss by one of two impedance-space losses: the squared error in (R, X) (z-raw), or the squared error in $(\log R, X)$ (z-log). The $1{,}000$-epoch z-log and z-raw runs give $\mathrm{MAE}(S)$ values of 0.300 and 0.321, respectively, against 0.00648 for the epoch- and seed-matched S-loss run.

E.3 Separate prediction of Q

Direct readout prediction. A second head predicts $\log|Q|$ while sharing the network backbone with the componentwise-S response model. Table 5 reports relative Q error read from the response S (relQ_S) and from the additional head (relQ_H).

F Modal contractivity of mapped reflection coordinates

For each scalar reflection coordinate mapped by Eq. 9, positivity of the represented resistance gives a contractive reflection coefficient.

Proposition 1. *Each scalar reflection coordinate to which $\Phi(s) = (\log R, X)$ is applied satisfies $|s| < 1$.*

Proof. The representation gives $Z = R + jX$ with $R > 0$ and $Z_0 > 0$. Therefore, $|Z - Z_0|^2 - |Z + Z_0|^2 = (R - Z_0)^2 - (R + Z_0)^2 = -4RZ_0 < 0$, so $|Z - Z_0| < |Z + Z_0|$ and $|s| < 1$. □

The scalar map is applied only to the reflection coordinates specified in §3 and Appendix A. The result therefore constrains only those represented reflection modes.

G Implementation details

Device geometry. Each device is an octagonal winding on the thick top metals, enclosed by a rectangular ground ring. The inner radius is r_{in}; for transformers, $r_{\mathrm{in,s}}$ denotes the secondary inner radius. The angles subtended at the center by the top and side edges are θ_1 and θ_2. The transformer here contains a one-turn primary and a one-turn secondary, each with a center tap to the other side of the ground ring. Table 6 lists the geometry parameters.

Architecture. Each geometry variable is standardized to zero mean and unit variance using training-split statistics. A two-hidden-layer encoder of width 512 maps the geometry vector to 16 latent tokens. For frequency $f \in [0, 200]\,\mathrm{GHz}$, let $\bar{f} = f/(200\,\mathrm{GHz})$. Its encoding is $[\bar{f}, \{\sin(\ell\pi\bar{f}), \cos(\ell\pi\bar{f})\}_{\ell=1}^{24}]$, followed by a linear projection. The decoder has four pre-normalized transformer layers with $d_{\mathrm{model}} = 256$, eight heads, feed-forward width 1024, and dropout 0.1. In each layer, the frequency tokens first self-attend over the queried frequency set, then serve as queries in cross-attention whose keys and values are the geometry tokens, and finally pass through the feed-forward block; each sublayer has a residual connection. The two-port models use three

Table 6: Geometry parameters of the three RF device templates. Lengths in μm, angles in degrees.

parameter	meaning	range
Differential inductor		
N	turns	1–4
r_{in}	inner radius	12–100
W	trace width	1–20
$\theta_1,\ \theta_2$	top and side angles	20–70
θ_3	feed-opening angle	5–$(\theta_2 - 10)$
L_{feed}	feed length	5–30
L_{space}	coil-to-ring clearance	0–30
W_{sub}	ring width	5–30
s	gap between turns	5–20
Transformer		
r_{in}	inner radius, primary	20–200
$r_{\mathrm{in,s}}$	inner radius, secondary	$0.8r_{\mathrm{in}}$–$1.2r_{\mathrm{in}}$
$W,\ W_{\mathrm{s}}$	trace width, primary and secondary	1–20
$\theta_1,\ \theta_2$	top and side angles	20–70
θ_3	feed-opening angle	5–$(\theta_2 - 10)$
$L_{\mathrm{feed}},\ L_{\mathrm{feed,s}}$	feed length, primary and secondary	10–r_{in}
$W_{\mathrm{feed}},\ W_{\mathrm{feed,s}},\ W_{\mathrm{ct}},\ W_{\mathrm{ct,s}}$	feed and center-tap widths	1–20
d	lateral offset of the secondary	-5–r_{in}
L_{space}	coil-to-ring clearance	10–r_{in}
W_{sub}	ring width	10–r_{in}
Single-ended spiral inductor		
N	turns	1–4
r_{in}	inner radius	10–150
W	trace width	3–15
$\theta_1,\ \theta_2$	top and side angles	20–70
$L_{\mathrm{r}},\ L_{\mathrm{t}},\ L_{\mathrm{l}},\ L_{\mathrm{b}}$	coil-to-ring clearance, per side	10–100
s	gap between turns	2–30

one-hidden-layer heads of width 128, while the six-port model uses 21; each head emits two real channels. These are grouped as $(\mathrm{Re}, \mathrm{Im})$ pairs for the componentwise-S baseline; for a coordinate model, the channels are interpreted in that model's coordinates and inverted to physical S before the loss. Appendix D compares the two-port network with a parameter-matched fixed-grid residual network.

Training. The main RF representation comparisons use AdamW (Kingma & Ba, 2015; Loshchilov & Hutter, 2019) with a learning rate of 2×10^{-3}, weight decay of 3×10^{-4}, a batch size of 1024 devices, and a 1,000-epoch budget. The learning rate warms linearly from 0.1 of its target value for ten epochs and then follows cosine decay. At each update, one set of 96 distinct frequencies is sampled without replacement and shared by all devices in the batch, yielding up to 1024×96 device–frequency evaluations. After each epoch, validation MAE is evaluated at all 201 frequencies and its running estimate is updated as $e_f^{(t)} = 0.7e_f^{(t-1)} + 0.3\,\mathrm{MAE}_f^{(t)}$. Sampling probabilities are proportional to $0.15 + \tilde{e}_f^2$, where $\tilde{e}_f = e_f/(F^{-1}\sum_f e_f)$; the same probability vector is used for the next epoch. Training uses early-stopping patience of 300, and all reported evaluations restore the checkpoint with the lowest full-frequency validation MSE rather than the final-epoch weights. Unless a section states otherwise, runs use training seeds 42, 1, and 2 with the data split fixed by seed 42.

Output and loss. In the primary representation comparison, the output map is inverted before the loss and the reconstructed physical response is compared with the same mean-squared S-space objective. Thus all response and circuit quantities are computed from physical scattering responses. The loss-modifying controls in Appendix E are exceptions and are identified there.

Compute and data. A run uses a 10 GB slice (one-seventh) of a single A100 GPU and takes from minutes to about an hour. Ground-truth responses are generated by a commercial electromagnetic solver on a 22 nm FD-SOI (22FDX) stack.

H OPEN-PDK REPLICATION

Process and solver. The replication uses the IHP SG13G2 open PDK (IHP GmbH, 2026), a 130 nm SiGe BiCMOS process with two thick top metals (2 and 3 μm) over a 2 S/m substrate, and the open finite-element solver Palace (Amazon Web Services, 2026) with second-order elements and an adaptive fast frequency sweep. Single-ended spiral inductors are drawn by the same ten-parameter generator as the 22FDX template and simulated at 201 frequencies from 0.1 to 200 GHz, as the solver cannot evaluate DC frequency. We generate 9,000 devices, split 7,200/900/900. The training setup is kept the same.

Table 7: Open-PDK replication of the single-ended spiral inductor rows of Table 1 (IHP SG13G2, Palace, 9,000 devices). Non-passive denotes the fraction of test points at which the reconstructed one-port has $|\hat{s}_{\mathrm{in}}| > 1$.

output representation	$\mathrm{MAE}(S)\downarrow$	$\mathrm{relQ}\downarrow$	non-passive
componentwise S	$\mathbf{0.02252}$ ± 0.00082	0.679 ± 0.089	0.22%
proposed	0.02311 ± 0.00031	$\mathbf{0.183}$ ± 0.016	0

Result. The representation reduces relative Q error by $3.7\times$ at a 2.6% change in $\mathrm{MAE}(S)$, with no Q supervision, and the reconstructed one-port is passive at every test point. The size of the reduction follows the componentwise amplification of Eq. 8, which grows with Q: the SG13G2 test set reaches a 90th-percentile in-band Q of 13.6 against 16.9 on 22FDX, and the reduction is $3.7\times$ against $6.8\times$.

I CIRCUIT-LEVEL EVALUATION

Table 3 pools design-level outcomes across the three training seeds. The test-device split and the 1,000 source-impedance targets are fixed across seeds.

LC-tank phase-noise evaluation. For each test differential inductor, the design frequency is $f_{\mathrm{d}} = 0.5 f_{\mathrm{SRF}}$. The surrogate sets an ideal lossless capacitor from its predicted reactance, $C = 1/(\omega_{\mathrm{d}} \hat{X}(f_{\mathrm{d}}))$, and predicts phase noise using $\hat{Q}(f_{\mathrm{d}})$. With this capacitor fixed, the field-solver response determines the realized resonance f_{osc} and $Q(f_{\mathrm{osc}})$. We use the Leeson $1/f^2$ expression at a 1 MHz offset with noise factor $F = 2$, signal power 1 mW, and temperature 300 K. Every test device is evaluated for every seed. Positive phase-noise violation means that the prediction is optimistic.

Transformer-library selection. The circuit consists of a differential source, a primary shunt capacitor, the six-port transformer, a secondary shunt capacitor, and a 50-Ω single-ended load. Ports (p_+, p_-) form the differential input and p_{CT} is grounded; s_+ drives the load, s_- is grounded, and s_{CT} is left open.

At 40 GHz, the 1,000 source impedances $Z_{\mathrm{src}} = R_{\mathrm{src}} + jX_{\mathrm{src}}$ form a stratified sample over $R_{\mathrm{src}} \in [10, 80]\,\Omega$ and $X_{\mathrm{src}} \in [-60, 20]\,\Omega$. For each target, the surrogate enumerates every test transformer with both shunt capacitors independently varied from 0 to 600 fF in 5 fF steps. It selects the candidate with the minimum predicted insertion loss among those satisfying the predicted $|S_{11}| < -10$ dB at the source reference plane.

The selected candidate is then evaluated with the field-solver response. The optimality gap is its realized insertion loss minus that of the best solver-feasible candidate in the same library. A specification miss occurs when its solver-evaluated $|S_{11}|$ does not satisfy the predicted -10 dB constraint.

Filtered baseline. The filtered componentwise-S baseline additionally requires the predicted differential-to-single-ended 2×2 scattering matrix at 40 GHz to satisfy $\sigma_{\max} \leq 1$.

J METASURFACE PROTOCOL

Simulation and data. The RCWA implementation returns magnetic-field modal amplitudes for unit incident magnetic-field amplitude. For incidence from air into a substrate of refractive index n_{sub}, the transmitted power is $\|t\|^2/n_{\mathrm{sub}}$; hence $\tilde{t} = t/\sqrt{n_{\mathrm{sub}}}$ in Eq. 11. A lossless structure gives unit reflected-plus-transmitted power to within 10^{-12}. The simulation retains 101 harmonics and uses a 64×64 grid in the patterned layer.

The unit cell is an elliptical dielectric pillar on an index-1.5 substrate. Its period, semi-axes, rotation, thickness, relative permittivity, and extinction coefficient κ are sampled from $[0.40, 0.65]\,\mu\mathrm{m}$, $[0.12, 0.42]\,\mu\mathrm{m}$, $[0^\circ, 180^\circ]$, $[0.15, 0.55]\,\mu\mathrm{m}$, $[2.5, 8.0]$, and $\log_{10}\kappa \in [-1.9, -0.5]$, respectively. Responses are evaluated over 1.45–$1.65\,\mu\mathrm{m}$.

Coordinate map. Let $\bar{v} = \mathrm{vec}_{\mathbb{R}}(v) = (\mathrm{Re}\, v, \mathrm{Im}\, v) \in \mathbb{R}^8$ and $d = U\bar{v}/\|\bar{v}\|_2$. The orthogonal matrix $U \in \mathbb{R}^{8\times 8}$ maps the normalized mean training-set direction to e_1. The directional coordinates are $w = d_{2:8}/(1 + d_1)$, with inverse

$$d(w) = \left(\frac{1 - \|w\|^2}{1 + \|w\|^2}, \frac{2w}{1 + \|w\|^2}\right), \qquad \bar{v} = U^\top d(w)\sqrt{1 - e^{-\mu}}, \qquad v = \mathrm{vec}_{\mathbb{R}}^{-1}(\bar{v}).$$

We construct U with a Householder reflection. A positive numerical floor is applied to μ before evaluating the square root. The inverse map is applied before the squared response loss.

Training and metrics. The experiment uses 48 wavelengths, $28{,}600$ training structures, and a separate test set. Each four-layer width-512 multilayer perceptron has $989{,}056$ parameters. We use Adam with learning rate 10^{-3}, batch size 256, cosine decay, $1{,}000$ epochs, and seeds 1, 2, and 42. The coordinate maps and response loss use float64 arithmetic. Relative absorptance error is $\mathrm{median}_{i,\lambda}\, |\hat{A}_{i\lambda} - A_{i\lambda}|/|A_{i\lambda}|$. Relative response-norm error is $\mathrm{median}_{i,\lambda}\, |\|\hat{v}_{i\lambda}\|_2 - \|v_{i\lambda}\|_2|/\|v_{i\lambda}\|_2$. MAE averages absolute error over the eight real response components.

K PER-DEVICE BEHAVIOR

Figure 5 shows predicted and solver responses for two devices from each of the three device templates. Each device is drawn as three stacked panels: its layout, the reflection magnitude, and the extracted quality factor, over the evaluated band of that device.

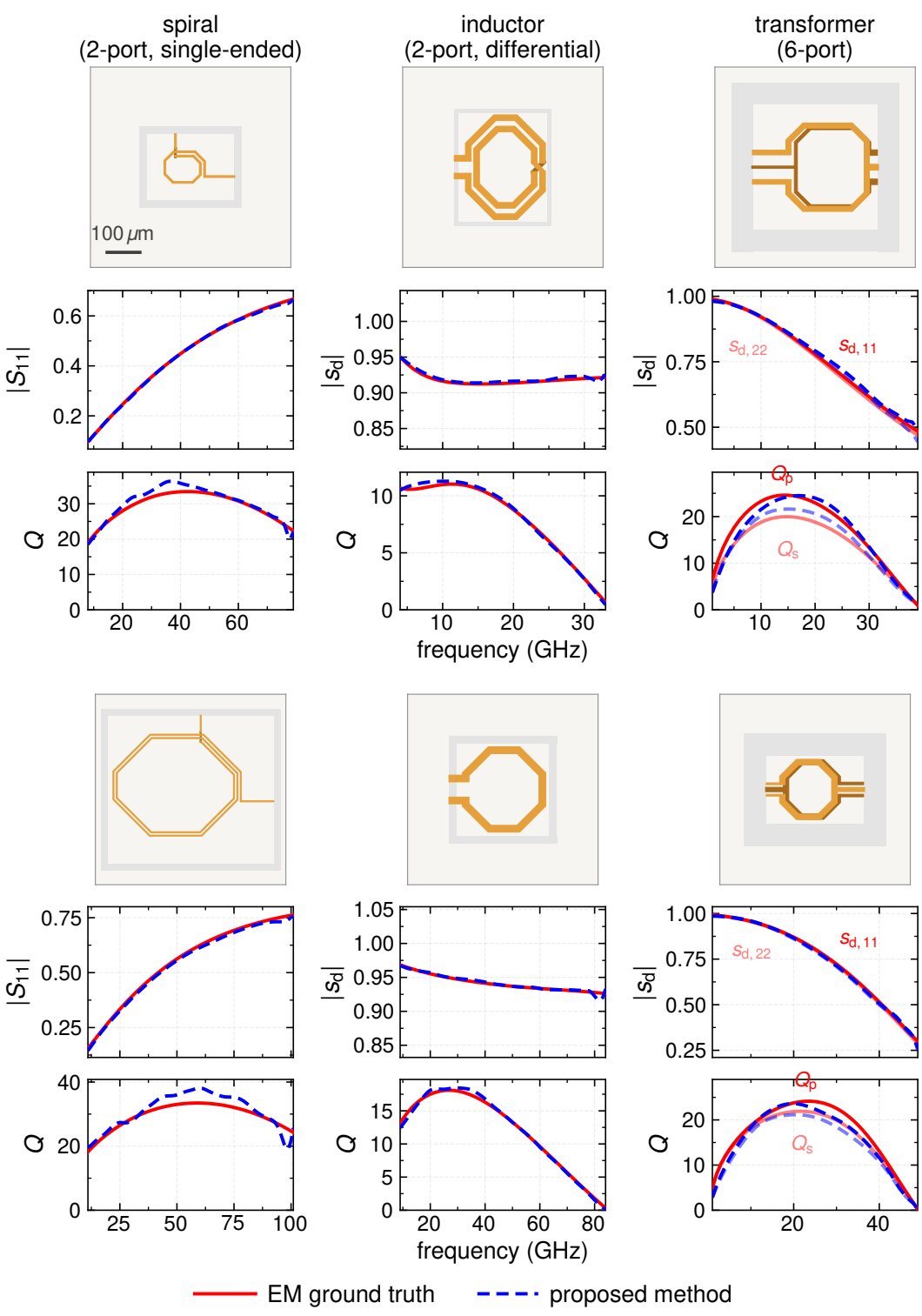

Figure 5: Predicted and solver responses for two devices from each template. Columns: single-ended spiral inductor (2 ports), differential inductor (2 ports), and transformer (6 ports); for the transformer both differential reflections and both quality factors are shown. Each device occupies three panels: its layout, the response magnitude, and the extracted quality factor. All six layouts share one scale, set by the scale bar in the first panel. Red is the electromagnetic ground truth and blue is the model trained in $(\log R, X)$ coordinates. Single training seed.