CONVERGENCE IS NOT ACCURACY: TRANSMISSION CONDITIONS AND NEURAL TRAINING IN FEM–PINN COUPLING

Anonymous authors
Paper under double-blind review

ABSTRACT

Hybrid solvers can compute one region of a domain with the finite-element method (FEM) and a neighboring region with a physics-informed neural network (PINN). The two solvers exchange boundary values and fluxes across their shared interface until the exchange settles. We show that when the PINN is inexact, the exchange can settle reliably on the wrong answer. For the classical Robin–Neumann exchange we prove that the Robin weight decides *whether and how fast* the exchange settles, while the *answer it settles on* is fixed by the PINN's interface response: for a fixed learned solver, every convergent Robin weight produces the same biased solution, with a closed-form error. Experiments on two-dimensional diffusion problems confirm this prediction to within 3×10^{-10}. Couplings that pass the same stopping test differ by more than an order of magnitude in error, and enlarging the learned solver's feature space removes most of the gap. Classical solves stopped at a loose tolerance show the same premature stopping. Frequency probes trace the PINN's response error to both its features and its finite training budget, and the budget also changes how well the training loss tracks the coupled error. Hybrid couplings should therefore report the exchanged response and the physical error, not only the iteration progress.

1 INTRODUCTION

Physics-informed neural networks (PINNs) fit a network to a PDE by penalizing its residual at sample points (Raissi et al., 2019). The finite-element method (FEM) solves the same PDE on a mesh with mature error control. Coupling the two is attractive when learning or measured data is useful in only part of a domain: FEM keeps its established discretization where it works well, and the PINN represents the neighboring region (Snyder et al., 2023; Shukla et al., 2025; Sobh et al., 2025). Figure 1(a) shows the configuration used throughout this paper.

The two solvers are joined by repeatedly exchanging data across the interface Σ, as in classical domain decomposition (Quarteroni & Valli, 1999; Gander, 2006). In the Robin–Neumann exchange of Figure 1(b), FEM receives a mixture of boundary value and flux weighted by a parameter $\alpha > 0$, and the PINN receives a flux. A well-chosen α turns an unstable exchange into a stable one. This remedy comes from fluid–structure interaction (Causin et al., 2005; Badia et al., 2008; Fernández et al., 2015), where it prevents the added-mass instability. Once the updates are small, the exchange is declared converged.

A PINN, however, is not an exact subdomain solver. What it returns for a given boundary problem depends on its features, its optimizer and training budget, and whether its weights are reused between exchanges. Classical subdomain solves can also be inexact, and inexact fixed-point theory already accounts for them (Birken, 2015). This paper asks what the inexactness of a *learned* solver does to the coupled answer, and which of its effects are specific to learning.

The answer turns on the difference between convergence and accuracy. A scalar example shows it. Suppose the FEM region responds to a boundary value with flux 1 per unit value, the PINN region with 8, and the correct interface value is 1. If the PINN captures only a quarter of its region's true response, the Robin–Neumann exchange with $\alpha = 2$ settles in a single step—at 3 (Appendix B).

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.

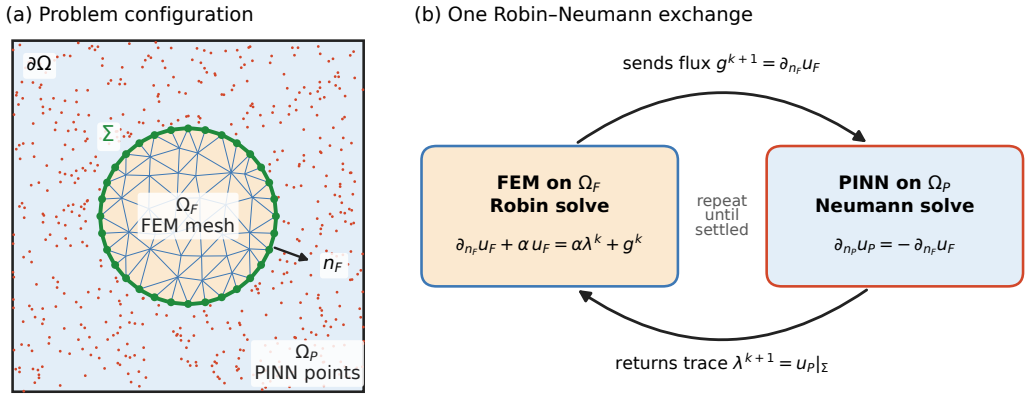

Figure 1: **The coupled problem.** (a) The square Ω is split along the circle Σ into a disc Ω_F, discretized by finite elements, and the surrounding region Ω_P, represented by a PINN trained on scattered collocation points. Green dots mark the shared interface nodes; n_F is the outward normal of Ω_F. (b) The Robin–Neumann exchange. FEM solves a Robin problem with data from the previous exchange and sends its flux to the PINN; the PINN solves a Neumann problem with that flux and returns its boundary value λ^{k+1}. The exchange repeats until successive boundary values stop changing.

Every other convergent α settles at the same wrong value. The update, the usual stopping signal, is exactly zero.

Contributions.

1. **A theorem for the Robin–Neumann exchange with a learned Neumann solver (Section 3).** Written in terms of the Robin data, the error obeys an exact recursion whose forcing is the learned solver's error at the correct flux. For a fixed learned solver, the limit is independent of the Robin weight and has a closed form.
2. **Verification on PDE solves (Section 4.1).** The exact FEM–FEM exchange rate is predicted to machine precision. Forty-five couplings with fixed learned solvers reach the predicted biased limit at every Robin weight. Retrained PINNs show the same insensitivity of error to the weight.
3. **Settled but wrong (Section 4.2).** On problems with prescribed interface frequencies, all fixed-feature couplings pass the stopping test, yet their errors span more than an order of magnitude. Classical solves stopped at a loose tolerance show the same premature stopping.
4. **What limits the response, and what the loss says about it (Sections 4.3–4.4).** Both the learned features and the finite optimization budget attenuate the transmitted flux. The training budget and weight-reuse policy change how well the training loss tracks the coupled error.

Scope. Manufactured solutions and exact discrete references make the interface error measurable. We claim no speedup over FEM, no universally optimal Robin weight, and no validated three-dimensional or multiphysics behavior.

2 The Coupled Problem and the Robin–Neumann Exchange

Consider $-\nabla \cdot (\kappa \nabla u) = f$ on the regions Ω_F and Ω_P of Figure 1, with prescribed data on $\partial\Omega$. On the interface Σ the solution must be continuous and its flux must balance:

$$u_F|_\Sigma = u_P|_\Sigma = \lambda, \qquad \kappa_F \partial_{n_F} u_F + \kappa_P \partial_{n_P} u_P = 0. \tag{1}$$

Here λ is the interface value, or *trace*, and n_i is the outward normal of region i. Each region, solved with boundary value λ, produces an outward flux that is affine in λ:

$$q_F(\lambda) = S_F\lambda - \chi_F, \qquad q_P(\lambda) = S_P\lambda - \chi_P, \qquad (S_F + S_P)\lambda^\star = \chi_F + \chi_P. \tag{2}$$

The matrix S_i is the region's Dirichlet-to-Neumann map: it says how much flux leaves the region when its boundary value changes. The offset χ_i accounts for forcing and exterior data. The coupled solution is the trace $\lambda^\star$ at which the fluxes balance. We measure traces in the interface mass norm $\|v\|_\Sigma^2 = v^\top M_\Sigma v$, where M_Σ is the interface mass matrix, a discrete boundary L^2 norm, and write $q^\star = q_F(\lambda^\star)$ for the correct interface flux.

The Robin–Neumann exchange. Each exchange passes one number per interface node in each direction (Figure 1b). With $\mu^k = \alpha\lambda^k + g^k$ denoting the *Robin data* and $A_\alpha = (\alpha I + S_F)^{-1}$:

1. *FEM Robin solve.* Solve on Ω_F with $\partial_{n_F} u_F + \alpha u_F = \mu^k$ on Σ. Its trace is $x^k = A_\alpha(\mu^k + \chi_F)$ and its outward flux is $q^k = S_F x^k - \chi_F$.
2. *PINN Neumann solve.* Train the PINN on Ω_P with outward flux $-q^k$ on Σ. Its trace is $\lambda^{k+1} = N_k(q^k)$.
3. *Update.* Set $g^{k+1} = q^k$, the flux just imposed on the PINN, so that $\mu^{k+1} = \alpha\lambda^{k+1} + q^k$.

For an exact Neumann solve, $N(q) = S_P^{-1}(\chi_P - q)$, and the trace error is multiplied at each exchange by

$$T = -S_P^{-1} S_F \, A_\alpha(\alpha I - S_P). \tag{3}$$

Two limits explain the role of α. As $\alpha \to 0$, $T \to I$ and the exchange stagnates. As $\alpha \to \infty$, $T \to -S_P^{-1}S_F$, the Dirichlet–Neumann exchange, which diverges when Ω_F responds more strongly than Ω_P. The Robin weight therefore has a useful band between these limits. Appendix A gives its modal form and the discrete conventions.

The learned Neumann solver. A trained PINN replaces the exact map N by the map N_k produced by training at exchange k. We describe it near the correct flux by two quantities:

$$N_k(q) = \lambda^\star + \eta_k - L_k(q - q^\star). \tag{4}$$

The *response error* $\eta_k = N_k(q^\star) - \lambda^\star$ is what the PINN gets wrong when it is handed the correct flux. The *sensitivity* L_k describes how its returned trace changes with the flux. An exact solver has $\eta = 0$ and $L = S_P^{-1}$. Both quantities are conditional on the training state and random draws at exchange k. Model (4) must hold on a set containing both the visited flux and $q^\star$; a fit at the visited flux alone determines neither.

3 Convergence decides speed, the learned response decides the answer

The following result specializes inexact fixed-point analysis (Birken, 2015) to the Robin–Neumann exchange. Tracking the Robin data rather than the trace avoids inverting the learned sensitivity L_k, which can be nearly singular for a learned solver.

Theorem 1 (Robin–Neumann exchange with a learned Neumann solver)**.** *Let S_F be symmetric positive semidefinite with smallest eigenvalue $s_F^{\min} \geq 0$, let $\alpha > 0$, and let the learned solver satisfy (4). Write $\delta^k = \mu^k - \mu^\star$ with $\mu^\star = \alpha\lambda^\star + q^\star$. Then*

$$\delta^{k+1} = M_k\,\delta^k + \alpha\,\eta_k, \qquad M_k = (I - \alpha L_k)\,S_F\,A_\alpha, \qquad x^k - \lambda^\star = A_\alpha\delta^k. \tag{5}$$

(a) ***Accumulated error.*** *If $\|M_k\| \leq \gamma < 1$ and $\|\eta_k\| \leq \bar\eta$ for all k in the mass norm, then the FEM trace satisfies*

$$\|x^k - \lambda^\star\| \leq \frac{1}{\alpha + s_F^{\min}}\Big(\underbrace{\gamma^k\|\delta^0\|}_{\text{decaying start}} + \underbrace{\alpha\,\bar\eta\,\frac{1-\gamma^k}{1-\gamma}}_{\text{accumulated response error}} \Big). \tag{6}$$

(b) ***Fixed learned solver.*** *If $L_k = L$ and $\eta_k = \eta$ for all k and the spectral radius of $M = (I - \alpha L)S_F A_\alpha$ is below one, then the FEM trace x^k and the PINN trace λ^k converge to the same limit $\widehat{\lambda}$, given by*

$$(I + LS_F)(\widehat{\lambda} - \lambda^\star) = \eta. \tag{7}$$

The limit does not depend on α. If L is invertible, then $\widehat{S}_P = L^{-1}$ is the learned region's effective Dirichlet-to-Neumann map and $\widehat{\lambda} - \lambda^\star = (S_F + \widehat{S}_P)^{-1}\widehat{S}_P\,\eta$.

Proof. The FEM flux error is $q^k - q^\star = S_F(x^k - \lambda^\star) = S_F A_\alpha \delta^k$. By (4), $\lambda^{k+1} - \lambda^\star = \eta_k - L_k(q^k - q^\star)$. Substituting both into $\delta^{k+1} = \alpha(\lambda^{k+1} - \lambda^\star) + (q^k - q^\star)$ gives (5). Iterating its norm inequality and using $\|A_\alpha\|_\Sigma = (\alpha + s_F^{\min})^{-1}$ gives (6). For (b), a spectral radius below one makes the affine iteration converge to its unique fixed point. At that point, the Robin equation $(\alpha I + S_F)\widehat{x} = \widehat{\mu} + \chi_F$ with $\widehat{\mu} = \alpha\widehat{\lambda} + \widehat{q}$ and $\widehat{q} = S_F\widehat{x} - \chi_F$ gives $\widehat{x} = \widehat{\lambda}$. Then (4) gives $\widehat{\lambda} - \lambda^\star = \eta - LS_F(\widehat{\lambda} - \lambda^\star)$, which is (7). The matrix $I + LS_F$ is invertible because $I - M = \alpha(I + LS_F)A_\alpha$ is. □

What the theorem says. The Robin weight enters the error map M but not the limit (7). Choosing α well, relaxing the update, or accelerating it with a method that preserves fixed points changes how fast the exchange settles, but *it cannot correct a wrong learned response*. An exact solver ($\eta = 0$) gives the correct answer; a learned one gives the answer of a slightly different interface problem. In the scalar example of Section 1, $s_F = 1$, $L = 1/2$, and $\eta = 3$, so the limit is $1 + \eta/(1 + Ls_F) = 3$ for every convergent α. The rate and the answer are separate questions. A small update $\|\lambda^{k+1} - \lambda^k\|$ answers only the first.

The bound (6) has the familiar form of perturbed contraction: an initial error that decays and a response error that accumulates. Its hypotheses concern the learned solver's response, not its training loss; sampled residuals alone establish neither γ nor $\bar{\eta}$. A fixed M with spectral radius below one contracts in *some* norm, but not necessarily the mass norm, and retraining makes M_k change between exchanges. Appendix C discusses these points, a nonlinear version of (6), and a variant that feeds back the PINN's own measured flux. In that variant, the limit acquires a dependence on α through the PINN's flux mismatch.

4 Experiments

The base problem is Poisson on $(0, 1)^2$, split into a disc of radius 0.25 (Ω_F) and the surrounding region (Ω_P), as in Figure 1. Its manufactured solution is $u^\star = \sin(\pi x)\sin(\pi y)$. FEM uses P1 triangles with nominal size $h = 0.06$ and 32 interface nodes. We use two kinds of learned Neumann solver. A *fixed-feature* solver refits the linear output layer of a bank of random sine features for each supplied flux, using one fixed least-squares rule. It is affine and does not change between exchanges, which makes Theorem 1(b) directly testable. A *PINN* is a width-64, depth-5 tanh network retrained by Adam (learning rate 10^{-3}) at every exchange, keeping its weights between exchanges. Protocols are in Appendices D and F.

4.1 The theorem on actual PDE solves

Exact exchange. With FEM on both regions, actual subdomain sweeps reproduce the full map (3). Its predicted spectral radius matches the measured dominant rate to within 4×10^{-15} at nine Robin weights from 2 to 800. The rate is fastest near $\alpha = 50$ (0.621) and exceeds one at $\alpha = 800$ (1.088): too large a weight reproduces the unstable Dirichlet–Neumann limit.

Fixed learned solvers. Across five feature seeds, three feature counts $(32, 128, 512)$, and $\alpha \in \{10, 25, 100\}$, all 45 couplings converge from zero data; the largest spectral radius of M is 0.959. Each run stops at the limit predicted by (7), to within 2.3×10^{-10} relative error. Depending on the weight, a run takes between about 40 and 330 exchanges to get there, but the three weights arrive at the same answer (Figure 2a). Improving the solver reduces the error of that answer: the median relative bias falls from 2.9% with 32 features to 0.75% with 128 and 0.26% with 512 (Figure 2b). These biases are measured against the exact discrete FEM–FEM solution. With 512 features held fixed, jointly refining the mesh and interface ($h = 0.06 \to 0.015$) reduces the median bias from

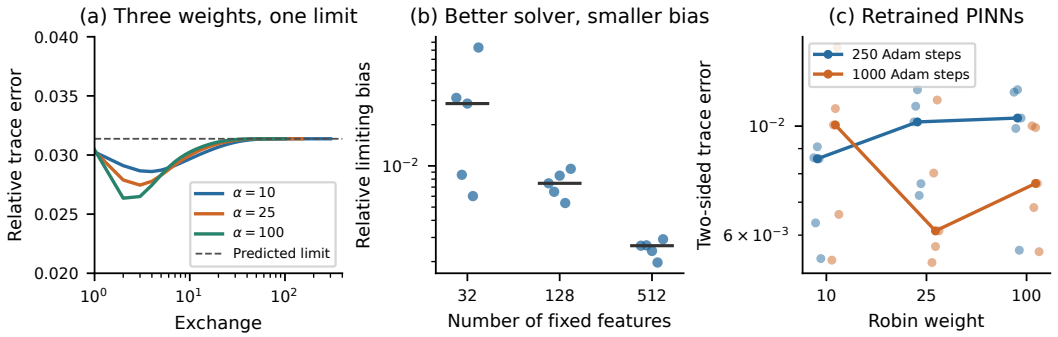

Figure 2: **The Robin weight sets the speed; the learned solver sets the answer.** (a) One fixed learned solver (32 features, seed 0) coupled to FEM with three Robin weights. All three converge, at different speeds, to the limit predicted by Theorem 1 (dashed). (b) More features give a better learned solver and a smaller limiting bias; dots are five seeds and bars their medians. (c) Retrained PINNs after 25 exchanges; dots are five seeds and lines their medians. The error is the larger of the FEM and PINN trace errors and hardly depends on the weight. Errors are relative to the exact discrete FEM–FEM trace.

2.6×10^{-3} to 1.1×10^{-4}. The coarse-grid plateau is therefore mostly discretization, not an irreducible neural floor.

Retrained PINNs. Retraining makes the learned map change between exchanges, so part (a) of the theorem applies only conditionally, and we measure outcomes instead. The grid has five seeds, $\alpha \in \{10, 25, 100\}$, and 250 or 1000 Adam steps per exchange (30 runs). Each run has 25 exchanges and starts from zero interface data without pretraining. All runs reach 1% trace error, after a median of 11 exchanges with 250 steps and 4–5 with 1000. After 25 exchanges the median trace error is between 0.61% and 0.69% in all six settings (Figure 2c). A tenfold change in α changes this error less than the spread across seeds, as the theorem predicts for the limit. The PINN's own Neumann-solve error at the flux it receives, 0.8–1.1% in median, has the same magnitude (Appendix D).

4.2 Settled but wrong: learned and classical inexact solves

The manufactured trace above is nearly constant, so it hardly tests high interface frequencies. We therefore add four problems whose exact interface trace is $\cos(j\theta)$: $j = 1, 4, 8$ at unit conductivity, and $j = 4$ with a tenfold conductivity in Ω_P. These use 128 interface nodes and $h = 0.015$. The Robin weight is chosen once per conductivity to minimize the exact exchange rate. Every run starts from zero and stops when successive traces differ by less than 10^{-9} in relative terms (Appendix E).

Learned solvers. Nested sets of 128, 512, and 1024 sine features share collocation points, loss weights, and the least-squares rule. All 60 couplings pass the stopping test, and each stops within 2.0×10^{-8} of its predicted limit. Nevertheless, at mode 8 the median global relative L^2 error is 3.53, 0.69, and 0.15 for the three feature counts (Figure 3a); FEM on both regions gives 0.036 on the same mesh. At mode 4, 128 features give an error of 5.1, while 512 and 1024 features reach the FEM level (0.0096 and 0.0094). Every one of these runs looks settled. Only the fitted function space determines whether the answer is right.

Classical solvers. To see which effects are specific to learning, we replace the learned solver by conjugate gradients on the FEM system of Ω_P. Six relative tolerances from 10^{-1} to 10^{-8} are each combined with a zero or warm (previous-solution) initial guess, giving 48 runs. All 48 pass the same stopping test. At loose tolerances they stop with trace errors up to 1.7%, and the initial guess changes that error. With a warm start the inner solver may make almost no progress, so the exchange stops after the minimum of ten exchanges (Figure 3b). At tolerance 10^{-8} every trace error is at most 1.0×10^{-8}. Premature stopping and dependence on history are therefore shared with classical inexact solves. The learned solvers add something classical ones lack: an error set by the fitted function space, which no inner tolerance controls.

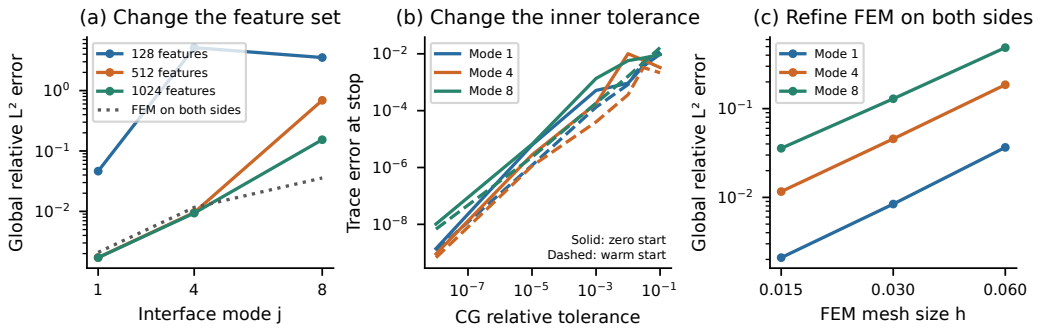

Figure 3: **Passing the stopping test does not certify accuracy.** (a) Global relative L^2 error of fixed-feature couplings that all pass the stopping test; medians over five seeds. Only the feature count changes within a seed. The dotted line is FEM on both regions. (b) Classical FEM solves on Ω_P by conjugate gradients, stopped at a relative tolerance: trace error when the exchange stops, from zero (solid) and warm (dashed) inner starts. (c) FEM on both regions under mesh refinement. All panels show unit conductivity; Appendix E lists all runs, including the contrast case.

An independent check. Residuals at held-out points can supplement the stopping test. Among the 60 settled feature runs, 31 have global error above 1%. A score formed from held-out PDE and exterior-boundary residuals, which uses no reference solution, exceeds 1% on 30 of them and on one of the 29 accurate runs (Appendix E). This is a post-hoc illustration on one benchmark, not a certified error estimator.

4.3 Both features and optimization limit the transmitted flux

Theorem 1 makes the learned solver's interface response, not its training loss, the quantity that decides the coupled error. To see what limits that response, we probe the PINN one boundary frequency at a time. We prescribe the boundary wave $v_j(\theta) = \cos(j\theta)$, train on that input, and project the returned outward flux onto the same wave:

$$\widehat{s}_j = \frac{v_j^\top M_\Sigma \left[\widehat{q}(v_j) - \widehat{q}(0)\right]}{v_j^\top M_\Sigma v_j}. \tag{8}$$

Dirichlet probes isolate one frequency cleanly, and a Neumann solve inverts the same trace–flux relation. On a disc of radius R the exact response is $s_j = j/R$: flux grows with frequency. A response that levels off instead attenuates the boundary signal. Appendix G validates the probe: five trainings recover the first three disc responses to within 1.1% median error. For Ω_P we compare with refined FEM responses (Appendix I). We summarize attenuation by the *cap index*, the first mode $j \geq 1$ with $|\widehat{s}_j| < 0.8\,|s_j|$; higher values mean that more frequencies are transmitted.

Two explanations compete for an attenuated response. The learned features may not span high-frequency boundary behavior, or the optimizer may not reach the best fit the features admit. Four comparisons over five seeds separate these. We train a multilayer perceptron (MLP) end to end with Adam, and we then freeze its learned features and refit only its linear output layer by least squares (LS). We also fit a fixed bank of 1600 random Fourier features by Adam and by LS, holding the sampled linear system fixed. Adam procedures use 2500 steps.

Figure 4 and Table 1 support both explanations. Refitting the learned output layer reduces low-frequency error but leaves the cap at $j = 2$ on every seed. For $j \geq 4$ its mean relative LS residual is about 0.98: the retained learned features barely fit these modes. This concerns the features produced by this training run, not the representational capacity of the architecture. For the fixed random features, switching from Adam to LS moves the mean cap index from 1.2 to 5.2, so those features support a much better fit than Adam reaches in 2500 steps. A coupling that transmits too few frequencies should therefore not be diagnosed as a capacity problem by default (Appendix I).

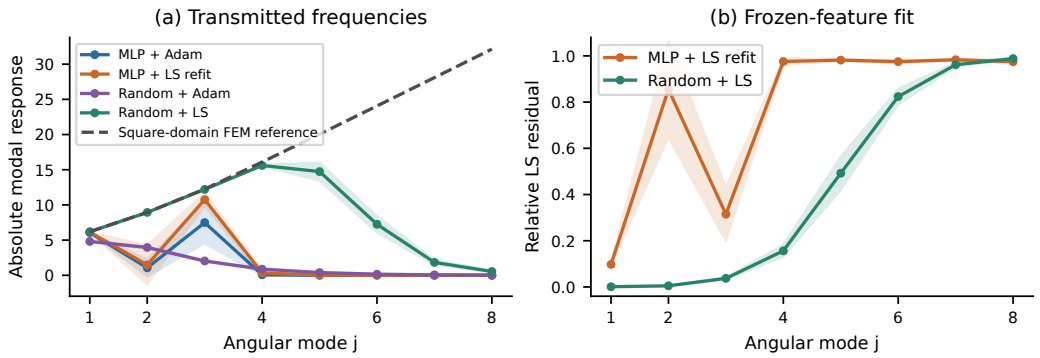

Figure 4: **Features and fitting both limit the transmitted flux.** (a) Mean absolute response to each boundary frequency over five seeds; bands show one standard deviation. The dashed curve is the refined FEM response of the same region. (b) Refitting frozen learned features leaves a large residual at high frequencies, whereas random Fourier features admit much better fits. Table 1 gives the cap indices.

Table 1: **Where each response first falls below** 80% **of the reference.** Mean $\pm$ standard deviation over five seeds. Low-band error averages the relative response errors of modes $j = 1, 2, 3$ against the refined FEM reference, whose last refinement changes them by at most 0.21%. Only the random-feature pair holds the sampled objective fixed; the learned-feature refit also changes it.

Features	Fitting procedure	Cap index	Low-band error (%)
Learned MLP	End-to-end Adam	2.0 ± 0.0	45.80 ± 7.91
Frozen learned MLP	Additional LS head refit	2.0 ± 0.0	32.53 ± 10.39
Fixed random Fourier	Adam head	1.2 ± 0.4	53.81 ± 0.81
Fixed random Fourier	LS head	5.2 ± 0.4	0.15 ± 0.02

4.4 Training budget and weight reuse change what the loss says

Each exchange changes the PINN's boundary data. We either keep the network weights between exchanges (warm starts) or reset them (cold restarts); both use a fresh Adam optimizer each time. In this study the PINN occupies the disc and FEM the surrounding region, with $\alpha = 10.313$ and sixteen exchanges. Warm starts cover nine budgets from 10 to 4000 Adam steps per exchange over twelve seeds (108 runs). Cold restarts cover 250, 1000, and 4000 steps with the same seeds (36 runs). Each run contributes its median trace error and the square root of its median training loss over the last six exchanges.

Pooling the shared budgets gives $R^2 = 0.875$ for warm starts but only 0.198 for cold restarts. Removing each budget's mean log loss and log error instead gives partial $R^2 = 0.870$ and 0.760, with 95% seed-bootstrap intervals $[0.758, 0.934]$ and $[0.611, 0.843]$. Cold fits at the three individual budgets have $R^2 = 0.478$, 0.822, and 0.873 (Figure 5). Budget and restart policy thus change how the loss maps to the error, rather than removing the association. These descriptive fits neither certify accuracy nor predict performance on new seeds or budgets.

5 Related work

Domain decomposition and partitioned coupling. Substructuring supplies the interface operators, Robin transmission, and acceleration used here (Quarteroni & Valli, 1999; Gander, 2006; Degroote et al., 2009). In fluid–structure interaction, the choice of transmission condition decides added-mass stability (Causin et al., 2005; Badia et al., 2008), and Robin–Neumann splittings yield stable, loosely coupled schemes (Fernández & Landajuela, 2013; Fernández et al., 2015; 2016; Landajuela et al., 2017). Inexact fixed-point theory already separates outer progress from inner-solve error (Birken, 2015). We use that framework to examine the feature, fitting, and training-history dependence of learned responses, rather than treating inexactness as unique to neural solvers.

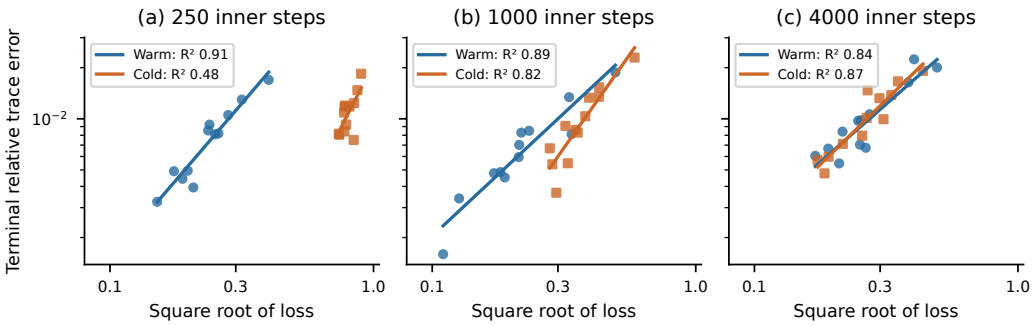

Figure 5: **Within a budget, the loss still tracks the error.** Each panel fixes the budget and shows twelve seeds per policy, with descriptive fits in log coordinates. The weak pooled cold-restart fit ($R^2 = 0.20$) hides a substantial association within each budget. Appendix H reports pooled and budget-adjusted fits with seed-bootstrap intervals.

Neural subdomains. Conservative and extended PINNs stitch networks across interfaces with residual and flux penalties (Jagtap et al., 2020; Jagtap & Karniadakis, 2020). Finite-basis PINNs use overlapping windowed networks (Moseley et al., 2023), and local extreme-learning machines fit linear heads per subdomain (Dong & Li, 2021); Heinlein et al. (2021) survey the area. These methods couple approximations of the same kind, usually through added loss terms. We study an exchange in which a trained solver and a classical discretization each solve their own boundary-value problem.

Neural–numerical coupling. Closest to our setting are Schwarz exchanges between PINNs and full-order models (Snyder et al., 2023), PINN–spectral-element coupling (Shukla et al., 2025), FEM–DeepONet interface coupling (Yin et al., 2022), variational interface condensation (Wu et al., 2026), FEM enforcement of PINN boundary conditions (Sobh et al., 2025), and FEM–PINN multifluid flow (Nohra & Dufour, 2024). Related formulations embed learned components inside FEM-discretized, PDE-constrained models (Mitusch et al., 2021). These works propose and demonstrate couplings; we ask what the exchange does when the learned side is inexact.

Inexact neural interface responses. That the transmitted flux is the weak point is recognized in learning-based substructuring. Sun et al. (2023) analyze the boundary-penalty error of a Dirichlet–Neumann learning algorithm, and Sun et al. (2024) introduce compensated extension operators to transmit flux across nonoverlapping interfaces. HINTS (Zhang et al., 2024) exploits the complementary spectral behavior of a learned operator and a relaxation method. Our contribution is diagnostic: we quantify how a given response error propagates into the coupled limit (7), and investigate what causes the attenuation such schemes are designed to tolerate. Spectral bias (Rahaman et al., 2019) and optimization failures (Wang et al., 2022; Krishnapriyan et al., 2021) motivate examining returned flux rather than loss alone. Fourier features with linear fitting (Tancik et al., 2020; Dong & Li, 2021; Chen et al., 2022) provide the controls of Section 4.3.

6 Discussion

Why this matters. Partitioned coupling is attractive for hybrid solvers because each solver can be built, trained, and checked on its own; the exchange between them is meant to be the easy part. With exact subdomain solves this is justified. A convergent exchange has only one fixed point, the coupled solution, so a small update is a reliable stand-in for accuracy, and the transmission condition matters only for speed and robustness. Our results show that this stand-in fails as soon as one subdomain solver is learned. The exchange still converges, sometimes faster than with an exact solver, and nothing in its iteration history shows that its fixed point has moved. Stability remedies imported from fluid–structure interaction solve the problem they were designed for, and only that problem.

Main conclusions.

1. **Convergence and accuracy are controlled by different things.** The Robin weight, relaxation, and acceleration that preserves fixed points act on the error map M, and therefore on the rate. The limit (7) depends only on the learned solver's sensitivity L and response error η. With a fixed learned solver, every convergent weight gives the same biased answer. In 45 PDE couplings this holds to within 2.3×10^{-10}. Retrained PINNs behave consistently with it: their final error (0.61–0.69%) hardly changes over a tenfold range of α.
2. **The error is set by the learned response, and it can be measured.** A better learned solver gives a smaller bias (2.9%, 0.75%, 0.26% with 32, 128, 512 features), and at 512 features most of what remains is FEM discretization error. When the interface data contains high frequencies, couplings that pass the same stopping test differ in error by more than an order of magnitude (3.53 versus 0.15 at mode 8).
3. **Stopping early is not specific to learning; the error floor is.** An inner conjugate-gradient solve stopped at a loose tolerance also settles early and depends on its starting guess, but its error disappears as the tolerance is tightened. A learned solver's error is fixed by its features and training, and no inner tolerance controls it.
4. **Features and optimization limit the response together.** Refitting a trained network's learned features by least squares does not raise its frequency cap, whereas the same refit on fixed random features does. A poor interface response should therefore not be attributed to network capacity by default.
5. **The training loss is an incomplete proxy.** Within a training budget the loss tracks the coupled error (partial $R^2 = 0.87$ warm, 0.76 cold), but pooled over budgets, cold restarts give $R^2 = 0.20$. A low loss does not replace a measurement of the interface response.

What to report. The theorem and experiments suggest a short checklist for hybrid couplings:

> **(1) The exchange:** which quantity each solver receives, the solve order, which solver returns the iterate, and which flux is fed back.
> **(2) The response, not the loss:** probe the learned solver on the boundary problem it actually solves; a low training loss certifies neither constant in Theorem 1.
> **(3) Rate and error separately:** a small update does not certify an accurate limit; report the error against an independent reference or residual check.
> **(4) Both stopping rules:** the first time a tolerance is reached and the error at a fixed budget can rank methods differently (Appendix K).
> **(5) Budget and history:** compare loss and error within a training budget, and state whether weights are reused.

Limitations. The evidence covers a circular interface with several manufactured trace spectra, a conductivity-contrast case, and a diffusion-dominated transport variant (Appendix M). The fixed-feature experiments test affine responses exactly; for retrained PINNs we measure outcomes but do not certify the uniform hypotheses of Theorem 1(a). The learned-feature refit changes its objective, so its conclusions are specific to the sampled system and rank threshold. Comparisons on broader geometries, at equal error and work, are needed. We establish no speedup over FEM.

Conclusion. When a learned model replaces a subdomain solver, the transmission condition still decides whether the exchange converges, but it no longer decides whether the answer is right. That is decided by the learned solver's response at the correct interface data, which depends on its features and training budget. A settled hybrid coupling should therefore be checked, not trusted: report the exchanged response, the iteration progress, the physical error, and the training protocol. Effort to improve such couplings is best spent on the learned solver's interface response, for example on features and training that transmit higher boundary frequencies, rather than on the transmission condition alone. The argument uses only the affine structure of the exchange and the response of the replaced solver, so we expect it to extend to other learned subdomain surrogates, such as neural operators; we have not tested this.

ETHICS STATEMENT

This work studies numerical algorithms on synthetic PDE benchmarks with manufactured solutions. It involves no human subjects, no personal or licensed data, and no deployment in a safety-critical setting. The main foreseeable risk is misplaced confidence in hybrid solvers: a coupling that appears to converge may be solving a biased interface problem. The purpose of this paper is to make that failure mode measurable rather than to advocate deployment.

REPRODUCIBILITY STATEMENT

Appendices A–C provide derivations and assumptions, and Appendices D–F specify every protocol. Appendix N maps each reported number to its record. The source bundle contains the records, the scripts that regenerate every figure, table, and summary from them, and runnable solver and training code for the Robin–Neumann verification, the controlled-mode study, the frequency probes, and the loss/restart study.

AI USE STATEMENT

Generative AI tools assisted research-code development, experiment planning and interpretation, mathematical formulation and proof drafting, critical audits of claims, literature discovery, manuscript organization and writing, and preparation of figures and LaTeX. All measured results come from executed numerical code. The authors take responsibility for all mathematical and empirical claims.

REFERENCES

Santiago Badia, Fabio Nobile, and Christian Vergara. Fluid–structure partitioned procedures based on Robin transmission conditions. *Journal of Computational Physics*, 227(14):7027–7051, 2008. doi: 10.1016/j.jcp.2008.04.006. URL https://doi.org/10.1016/j.jcp.2008.04.006.

Philipp Birken. Termination criteria for inexact fixed-point schemes. *Numerical Linear Algebra with Applications*, 22(4):702–716, 2015. doi: 10.1002/nla.1982. URL https://doi.org/10.1002/nla.1982.

Paola Causin, Jean-Frédéric Gerbeau, and Fabio Nobile. Added-mass effect in the design of partitioned algorithms for fluid–structure problems. *Computer Methods in Applied Mechanics and Engineering*, 194(42–44):4506–4527, 2005. doi: 10.1016/j.cma.2004.12.005. URL https://doi.org/10.1016/j.cma.2004.12.005.

Jingrun Chen, Xurong Chi, Weinan E, and Zhouwang Yang. Bridging traditional and machine learning-based algorithms for solving PDEs: The random feature method. *Journal of Machine Learning*, 1(3):268–298, 2022. doi: 10.4208/jml.220726. URL https://doi.org/10.4208/jml.220726.

Joris Degroote, Klaus-Jürgen Bathe, and Jan Vierendeels. Performance of a new partitioned procedure versus a monolithic procedure in fluid–structure interaction. *Computers & Structures*, 87(11–12):793–801, 2009. doi: 10.1016/j.compstruc.2008.11.013. URL https://doi.org/10.1016/j.compstruc.2008.11.013.

Suchuan Dong and Zongwei Li. Local extreme learning machines and domain decomposition for solving linear and nonlinear partial differential equations. *Computer Methods in Applied Mechanics and Engineering*, 387:114129, 2021. doi: 10.1016/j.cma.2021.114129. URL https://doi.org/10.1016/j.cma.2021.114129.

Miguel A. Fernández and Mikel Landajuela. A fully decoupled scheme for the interaction of a thin-walled structure with an incompressible fluid. *Comptes Rendus Mathématique*, 351(3–4):161–164, 2013. doi: 10.1016/j.crma.2013.02.015. URL https://doi.org/10.1016/j.crma.2013.02.015.

Miguel A. Fernández, Mikel Landajuela, and Marina Vidrascu. Fully decoupled time-marching schemes for incompressible fluid/thin-walled structure interaction. *Journal of Computational Physics*, 297:156–181, 2015. doi: 10.1016/j.jcp.2015.05.009. URL https://doi.org/10.1016/j.jcp.2015.05.009.

Miguel A. Fernández, Mikel Landajuela, Jimmy Mullaert, and Marina Vidrascu. Robin–Neumann schemes for incompressible fluid–structure interaction. In Thomas Dickopf, Martin J. Gander, Laurence Halpern, Rolf Krause, and Luca F. Pavarino (eds.), *Domain Decomposition Methods in Science and Engineering XXII*, Lecture Notes in Computational Science and Engineering, pp. 65–76. Springer, 2016. doi: 10.1007/978-3-319-18827-0_6. URL https://doi.org/10.1007/978-3-319-18827-0_6.

Martin J. Gander. Optimized Schwarz methods. *SIAM Journal on Numerical Analysis*, 44(2): 699–731, 2006. doi: 10.1137/S0036142903425409. URL https://doi.org/10.1137/S0036142903425409.

Alexander Heinlein, Axel Klawonn, Martin Lanser, and Janine Weber. Combining machine learning and domain decomposition methods for the solution of partial differential equations—a review. *GAMM-Mitteilungen*, 44(1):e202100001, 2021. doi: 10.1002/gamm.202100001. URL https://doi.org/10.1002/gamm.202100001.

Ameya D. Jagtap and George Em Karniadakis. Extended physics-informed neural networks (XPINNs): A generalized space-time domain decomposition based deep learning framework for nonlinear partial differential equations. *Communications in Computational Physics*, 28(5): 2002–2041, 2020. doi: 10.4208/cicp.OA-2020-0164. URL https://doi.org/10.4208/cicp.OA-2020-0164.

Ameya D. Jagtap, Ehsan Kharazmi, and George Em Karniadakis. Conservative physics-informed neural networks on discrete domains for conservation laws: Applications to forward and inverse problems. *Computer Methods in Applied Mechanics and Engineering*, 365:113028, 2020. doi: 10.1016/j.cma.2020.113028. URL https://doi.org/10.1016/j.cma.2020.113028.

Aditi S. Krishnapriyan, Amir Gholami, Shandian Zhe, Robert M. Kirby, and Michael W. Mahoney. Characterizing possible failure modes in physics-informed neural networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pp. 26548–26560, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/df438e5206f31600e6ae4af72f2725f1-Abstract.html.

Mikel Landajuela, Marina Vidrascu, Dominique Chapelle, and Miguel A. Fernández. Coupling schemes for the FSI forward prediction challenge: Comparative study and validation. *International Journal for Numerical Methods in Biomedical Engineering*, 33(4):e2813, 2017. doi: 10.1002/cnm.2813. URL https://doi.org/10.1002/cnm.2813.

Sebastian K. Mitusch, Simon W. Funke, and Miroslav Kuchta. Hybrid FEM–NN models: Combining artificial neural networks with the finite element method. *Journal of Computational Physics*, 446: 110651, 2021. doi: 10.1016/j.jcp.2021.110651. URL https://doi.org/10.1016/j.jcp.2021.110651.

Ben Moseley, Andrew Markham, and Tarje Nissen-Meyer. Finite basis physics-informed neural networks (FBPINNs): A scalable domain decomposition approach for solving differential equations. *Advances in Computational Mathematics*, 49(4):62, 2023. doi: 10.1007/s10444-023-10065-9. URL https://doi.org/10.1007/s10444-023-10065-9.

Michel Nohra and Steven Dufour. Coupling of the finite element method with physics informed neural networks for the multi-fluid flow problem. *arXiv preprint arXiv:2405.05371*, 2024. URL https://arxiv.org/abs/2405.05371.

Alfio Quarteroni and Alberto Valli. *Domain Decomposition Methods for Partial Differential Equations*. Oxford University Press, 1999. URL https://academic.oup.com/book/53865.

Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred A. Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks. In *Proceedings*

of the 36th International Conference on Machine Learning (ICML), volume 97 of *Proceedings of Machine Learning Research*, pp. 5301–5310, 2019. URL https://proceedings.mlr.press/v97/rahaman19a.html.

Maziar Raissi, Paris Perdikaris, and George E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019. doi: 10.1016/j.jcp.2018.10.045. URL https://doi.org/10.1016/j.jcp.2018.10.045.

Khemraj Shukla, Zongren Zou, Chi Hin Chan, Additi Pandey, Zhicheng Wang, and George Em Karniadakis. NeuroSEM: A hybrid framework for simulating multiphysics problems by coupling PINNs and spectral elements. *Computer Methods in Applied Mechanics and Engineering*, 433: 117498, 2025. doi: 10.1016/j.cma.2024.117498. URL https://arxiv.org/abs/2407.21217.

William Snyder, Irina Tezaur, and Christopher R. Wentland. Domain decomposition-based coupling of physics-informed neural networks via the Schwarz alternating method. *arXiv preprint arXiv:2311.00224*, 2023. URL https://arxiv.org/abs/2311.00224.

Nahil Sobh, Rini Jasmine Gladstone, and Hadi Meidani. PINN-FEM: A hybrid approach for enforcing Dirichlet boundary conditions in physics-informed neural networks. *arXiv preprint arXiv:2501.07765*, 2025. URL https://arxiv.org/abs/2501.07765.

Qi Sun, Xuejun Xu, and Haotian Yi. Dirichlet–Neumann learning algorithm for solving elliptic interface problems. *arXiv preprint arXiv:2301.07361*, 2023. URL https://arxiv.org/abs/2301.07361.

Qi Sun, Xuejun Xu, and Haotian Yi. Domain decomposition learning methods for solving elliptic problems. *SIAM Journal on Scientific Computing*, 46(4):A2445–A2474, 2024. doi: 10.1137/22M1515392. URL https://doi.org/10.1137/22M1515392.

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 (NeurIPS)*, volume 33, pp. 7537–7547, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/55053683268957697aa39fba6f231c68-Abstract.html.

Sifan Wang, Xinling Yu, and Paris Perdikaris. When and why PINNs fail to train: A neural tangent kernel perspective. *Journal of Computational Physics*, 449:110768, 2022. doi: 10.1016/j.jcp.2021.110768. URL https://doi.org/10.1016/j.jcp.2021.110768.

Chang Wu, Chang Liu, Yilin Guo, and Xu Guo. DFENN: A penalty-free variational framework coupling finite elements and neural networks via interface condensation. *Journal of the Mechanics and Physics of Solids*, 215:106703, 2026. doi: 10.1016/j.jmps.2026.106703. URL https://doi.org/10.1016/j.jmps.2026.106703.

Minglang Yin, Enrui Zhang, Yue Yu, and George Em Karniadakis. Interfacing finite elements with deep neural operators for fast multiscale modeling of mechanics problems. *Computer Methods in Applied Mechanics and Engineering*, 402:115027, 2022. doi: 10.1016/j.cma.2022.115027. URL https://arxiv.org/abs/2203.00003.

Enrui Zhang, Adar Kahana, Alena Kopaničáková, Eli Turkel, Rishikesh Ranade, Jay Pathak, and George Em Karniadakis. Blending neural operators and relaxation methods in PDE numerical solvers. *Nature Machine Intelligence*, 6(11):1303–1313, 2024. doi: 10.1038/s42256-024-00910-x. URL https://doi.org/10.1038/s42256-024-00910-x.

A EXCHANGE MAPS, SIGN CONVENTIONS, AND DISCRETE COORDINATES

Let $q_i = S_i\lambda - \chi_i$ be the flux of region i in its own outward normal. Flux balance is $q_F + q_P = 0$, and the exact target is $(S_F + S_P)\lambda^\star = \chi_F + \chi_P$. The linear part S_i is also called the Steklov–Poincaré operator. With these conventions every S_i is positive semidefinite, and no sign has to be attached to an otherwise positive Dirichlet-to-Neumann operator. For piecewise conductivity, fluxes are conormal fluxes $\kappa_i \partial_{n_i} u_i$.

Discrete coordinates. If K_i maps trace degrees of freedom to integrated reactions and M_Σ is the interface mass matrix, then pointwise fluxes are $q_{\mathrm{pw}} = M_\Sigma^{-1} q_{\mathrm{int}}$ and $S_i = M_\Sigma^{-1} K_i$. Under $\widetilde{\lambda} = M_\Sigma^{1/2}\lambda$ the operator becomes $\widetilde{S}_i = M_\Sigma^{-1/2} K_i M_\Sigma^{-1/2}$, which is symmetric; operator norms in the main text are taken in these mass-orthonormal coordinates. A weak Robin solve uses $K_F + \alpha M_\Sigma$, whose pointwise analogue is $S_F + \alpha I$. All three forms must represent the same α.

Two equivalent forms of the exact exchange. With an exact Neumann solve, the exchange of Section 2 maps traces by $T = -S_P^{-1} S_F A_\alpha (\alpha I - S_P)$ and Robin data by $M = (I - \alpha S_P^{-1}) S_F A_\alpha$. The second form is the case $L = S_P^{-1}$, $\eta = 0$ of Theorem 1. Because S_F and A_α commute, $M = S_P^{-1}(S_P - \alpha I) S_F A_\alpha$ and $T = S_P^{-1} S_F A_\alpha (S_P - \alpha I)$ have the same spectrum, so the two forms predict the same rate. The Robin-data form remains well defined when the learned sensitivity L is nearly singular, as it is for the fixed-feature solvers of Section 4.1 (Appendix D).

A.1 MODAL BAND OF THE ROBIN WEIGHT

When both exact operators act on the same orthonormal boundary modes with responses $s_F, s_P > 0$, each mode's error is multiplied at every exchange by

$$t = -\frac{s_F}{s_P}\,\frac{\alpha - s_P}{\alpha + s_F}. \tag{9}$$

The factor vanishes at $\alpha = s_P$, the modal analogue of the classical fluid–structure choice of the Robin weight. It has magnitude below one precisely when $\alpha(s_F - s_P) < 2 s_F s_P$. Hence every $\alpha > 0$ is admissible for modes with $s_F \le s_P$; otherwise the band is bounded above by $2 s_F s_P/(s_F - s_P)$. As $\alpha \to \infty$, $t \to -s_F/s_P$, the Dirichlet–Neumann factor, which exceeds one in magnitude when $s_F > s_P$. The admissible band is the intersection over modes. Discrete subdomains need not share eigenvectors, so our references use the full matrix spectrum. Nor should the pairing of modes be inferred from separately sorted eigenvalues when the operators do not commute.

A.2 NEUMANN PROBLEMS ON THE DISC

With exterior Dirichlet data, the Neumann problem on the surrounding region Ω_P of Figure 1 is well posed. A Neumann problem on the disc is not: constant traces produce zero flux, so the trace is determined only up to a constant. The loss/restart study of Section 4.4 places the PINN on the disc. Its driver fixes the constant by setting the mean of the returned PINN trace equal to the mean of the current FEM trace, and it feeds back the PINN's measured flux. This moving normalization is not the fixed projection of the algebra above, so that study is reported as an empirical result without applying Theorem 1 to it.

B THE SCALAR EXAMPLE

Take $s_F = 1$, $s_P = 8$, $\chi_F = 0$, and $\chi_P = 9$, so that $\lambda^\star = 1$. Suppose the learned Neumann solver behaves like a region that retains a fraction c of the true response: $N(q) = (9 - q)/(8c)$. Then $L = 1/(8c)$, $\eta = 8(1 - c)/(8c)$, the Robin-data multiplier is $(1 - \alpha/(8c))/(\alpha + 1)$, and (7) gives the limit $\widehat{\lambda} = 9/(1 + 8c)$. The limit does not involve α. At $c = 1/4$ and $\alpha = 2$ the multiplier is zero, so the exchange settles in one step at $\widehat{\lambda} = 3$. The exact exchange ($c = 1$) converges to 1 with multiplier $1/4$. Figure 6 plots these closed forms; they are analytic illustrations, separate from the measured results.

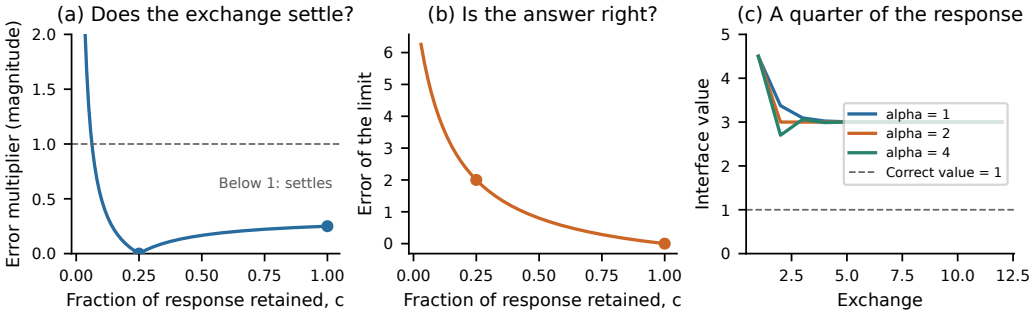

Figure 6: **An exchange can converge to the wrong answer.** Scalar example of Appendix B, in which the learned solver retains a fraction c of the true response. (a) At $\alpha = 2$ the exchange settles for all $c > 1/16$. (b) The error of its limit grows as c decreases. (c) At $c = 1/4$, three Robin weights all settle at 3 instead of the correct value 1.

C RETRAINING, NONLINEAR RESPONSES, AND FLUX FEEDBACK

Nonlinear responses. Theorem 1 does not require a global affine model. Conditional on the weights, optimizer state, and random draws at the start of exchange k, let $\Phi_k(\mu) = \alpha N_k(q(\mu)) + q(\mu)$ be the Robin-data update, where $q(\mu) = S_F A_\alpha(\mu + \chi_F) - \chi_F$. The correct data satisfy $\Phi_k(\mu^\star) - \mu^\star = \alpha\eta_k$. If $\|\Phi_k(v) - \Phi_k(w)\| \le \gamma\|v - w\|$ on a set containing μ^k and $\mu^\star$, then

$$\|\delta^{k+1}\| \le \gamma\|\delta^k\| + \alpha\|\eta_k\|,$$

and summing gives (6). Both evaluations must start from the same training state; a fit at the visited input alone establishes neither bound. Uniformly bounded, time-varying errors give an error neighborhood, not necessarily a convergent trajectory. Only a fixed affine solver gives the closed-form limit (7).

A common norm is required. Part (a) needs one norm in which every realized M_k contracts. The spectral radius of each frozen matrix is not enough for a time-varying product. For example,

$$B_1 = \begin{pmatrix} 0 & a \\ 0 & 0 \end{pmatrix}, \qquad B_2 = \begin{pmatrix} 0 & 0 \\ a & 0 \end{pmatrix} \tag{10}$$

are both nilpotent, but $B_1 B_2$ has eigenvalue $a^2 > 1$ when $a > 1$. For a single fixed M with spectral radius below one, a contracting norm always exists, for instance from the discrete Lyapunov equation $X - M^\top X M = I$, but it need not be the mass norm. For the fixed-feature solvers of Section 4.1, $\|M\|_\Sigma$ is between 1.0 and 3.0 although its spectral radius is at most 0.959. Converting between such a norm and the mass norm introduces equivalence constants into (6). Part (b) uses only the spectral radius.

Feeding back the measured flux. Step 3 of the exchange feeds back the flux imposed on the PINN. Some drivers instead feed back the PINN's measured flux, $g^{k+1} = q^k + \zeta_k$, where ζ_k is the PINN's flux mismatch. For an exact Neumann solve $\zeta_k = 0$, but a penalty-trained PINN need not reproduce its flux exactly. The recursion becomes $\delta^{k+1} = M_k\delta^k + \alpha\eta_k + \zeta_k$. At a fixed point with constant ζ, the Robin equation gives $\widehat{x} = \widehat{\lambda} + \zeta/\alpha$, and

$$(I + LS_F)(\widehat{\lambda} - \lambda^\star) = \eta - LS_F\,\zeta/\alpha. \tag{11}$$

The flux mismatch therefore adds an α-dependent term, which vanishes as $\alpha \to \infty$, and the FEM and PINN traces no longer coincide in the limit. The retrained PINN runs of Section 4.1 feed back the imposed flux and record the measured mismatch separately (Table 6). The rate sweeps, Schwarz comparison, and loss/restart study of Appendices J–L and Section 4.4 use drivers that feed back the measured flux.

From loss to response. A concrete bound on $\bar{\eta}$ from the training loss would additionally require well-posedness, boundary compatibility, regularity, and control of the population PDE and boundary residuals in suitable norms. Sampled squared residuals with penalty weights do not provide these by themselves. The budget regressions of Section 4.4 are descriptive and are not used to certify (6).

D Robin–Neumann verification: protocols and full results

All runs in this appendix use the square/disc geometry with 32 shared interface nodes, nominal mesh size $h = 0.06$, and $u^\star = \sin(\pi x)\sin(\pi y)$. FEM occupies the disc and performs the Robin solve; the learned solver occupies the surrounding region and performs the Neumann solve, which is well posed because of the exterior Dirichlet data. No mean normalization is used. The disc's positive-semidefinite response has a constant kernel, covered by Theorem 1 with $s_F^{\min} = 0$ and $\alpha > 0$.

Reference and sign checks. Each region's flux uses its own outward normal. Integrated FEM reactions are converted to pointwise fluxes with the shared interface mass matrix. Independent sparse Dirichlet solves supply the offsets χ_i, and the exact discrete reference is $(S_F + S_P)^{-1}(\chi_F + \chi_P)$. Its relative mass-norm distance from the manufactured trace is 3.43×10^{-3}. Errors measured against the discrete reference therefore include the difference between the continuous learned problem and the discrete FEM problem, including the polygonal approximation of the interface.

Exact exchange. Homogeneous sweeps with actual sparse solves, applied to every interface basis vector, reproduce the full map (3) to within 5×10^{-15}. Forced sweeps preserve the discrete fixed point. We also start a homogeneous error in the dominant eigenvector and measure 40 successive norm ratios, renormalizing after each sweep; the median of the last ten is the measured rate in Table 2.

Table 2: **Predicted and measured rates of the exact exchange** (FEM on both regions), at seven of the nine tested weights. The rate exceeds one at $\alpha = 800$.

α	$\rho(T)$	Measured rate
2	0.981143	0.981143
10	0.909065	0.909065
25	0.786968	0.786968
50	0.621426	0.621426
100	0.747322	0.747322
200	0.828375	0.828375
800	1.088349	1.088349

Fixed learned solvers. For each of five seeds, draw a bank of 512 sine features with frequencies $w \sim \mathcal{N}(0, 6^2 I)$ and phases uniform on $[0, 2\pi]$, plus a constant, and use nested subsets of 32, 128, and 512 features. Within a seed, the same 2048 interior points, 512 exterior boundary points, and 32 interface nodes are used for all feature counts and weights. The output coefficients are fit to the PDE residual, the interface Neumann condition (outward flux equal to the supplied flux), and the exterior Dirichlet condition, with mean-square weights $1, 100, 100$. Column equilibration and a fixed SVD cutoff of 10^{-10} define one least-squares rule. Refitting the head for each supplied flux yields a fixed affine map $\lambda = N(q)$, from which we read off L and η exactly. Its sensitivity L is severely ill conditioned (condition numbers between about 10^8 and 10^{17}), because the fitted solver returns almost no trace for high-frequency flux; this is why Theorem 1 is stated in terms of the Robin data.

Each run starts from zero trace and flux and executes an actual sparse FEM Robin solve and a learned Neumann solve at every exchange, for $\alpha \in \{10, 25, 100\}$. All 45 runs have spectral radius of M below 0.959 and converge; their limits agree with (7) to within 2.3×10^{-10} relative to the norm of the discrete reference. Runs stop at relative distance 10^{-10} from the prediction, or at an update below 10^{-11} with distance below 10^{-8}, with a cap of 5000 exchanges. These prediction-based stopping conditions are verification diagnostics, not deployable stopping rules.

Refinement. Holding the 512-feature draws and interior/exterior samples fixed, we refine the FEM mesh and interface jointly and recompute the limit of each seed at $\alpha = 25$. An actual Robin solve

Table 3: **A better learned solver gives a smaller limiting bias.** Five seeds; the bias does not depend on α. $\|\eta\|$ is the learned solver's trace error when given the exact discrete flux, relative to the reference trace. Percentages include continuous/discrete mismatch.

Features	Median bias (%)	Bias range (%)	Median $\|\eta\|$ (%)	Largest ρ
32	2.851	[0.601, 7.318]	1.109	0.958
128	0.746	[0.535, 0.952]	0.204	0.959
512	0.261	[0.197, 0.291]	0.195	0.959

and learned Neumann solve reproduce each predicted limit to within 10^{-9}. The median bias to the discrete reference falls with refinement, and the bias to the analytic trace tracks the FEM reference error (Table 4).

Table 4: **The coarse-grid bias is mostly discretization.** 512 fixed features, five seeds (medians), under joint mesh and interface refinement.

n_Σ	h	FEM reference error	Bias to discrete	Bias to analytic
32	0.060	0.003426	0.002612	0.003884
64	0.030	0.000980	0.000436	0.000860
128	0.015	0.000246	0.000107	0.000249

Retrained PINNs. The 30-run grid has five seeds, three weights $\alpha \in \{10, 25, 100\}$, and two budgets of 250 or 1000 Adam steps per exchange. Each run has 25 exchanges, 256 resampled interior points and 128 resampled exterior points per Adam step, a width-64/depth-5 tanh network, learning rate 10^{-3}, and boundary loss weights 100. Networks start fresh with zero interface trace and flux and no pretraining. They keep their weights between exchanges and use a fresh Adam optimizer each time. The returned-trace error uses the PINN trace; the two-sided error is the larger of the PINN and FEM trace errors. We also record the PINN's Neumann-solve error at the flux it receives, measured against the exact discrete Neumann solve, and the mismatch between the PINN's measured flux and the imposed flux at 256 held-out angular points. Wall times are recorded but not used for performance claims: several single-threaded CPU jobs ran concurrently.

Table 5: **Retrained PINN couplings.** Medians over five seeds after 25 exchanges. First hit counts the exchanges needed to reach 1% returned-trace error, with successes in parentheses.

Inner steps	α	Returned error	Two-sided error	Update	First hit
250	10	0.0064	0.0086	0.0035	11 (5/5)
250	25	0.0069	0.0102	0.0077	11 (5/5)
250	100	0.0067	0.0104	0.0076	11 (5/5)
1000	10	0.0066	0.0101	0.0086	5 (5/5)
1000	25	0.0061	0.0061	0.0071	4 (5/5)
1000	100	0.0068	0.0076	0.0074	4 (5/5)

Reproduction. The bundle contains all 47 JSON records (one exact-reference record, 15 fixed-feature records, one refinement record, and 30 retrained runs) with source hashes, together with analysis code that checks the complete grids and regenerates the figure and tables, and the runnable driver with its dependencies. A smoke option uses one seed, one weight, and three short exchanges; smoke results are excluded from all reported aggregates.

E Controlled interface modes and an inexact FEM comparison

The base manufactured trace is nearly constant: its constant component carries 99.9946% of the analytic squared boundary norm, computed on 4096 equally spaced angles. The problems of Section 4.2 instead impose an interface mode of unit amplitude. Let $z = (x - 0.5) + \mathrm{i}(y - 0.5)$,

Table 6: **Interface diagnostics of the retrained couplings.** Full five-seed range of the final returned-trace error, the median Neumann-solve error at the received flux (relative to the reference trace), and the median RMS mismatch between measured and imposed flux at 256 held-out points (absolute).

Inner steps	α	Error range	Neumann response error	Held-out flux mismatch
250	10	[0.0046, 0.0086]	0.0096	0.1264
250	25	[0.0056, 0.0110]	0.0100	0.1620
250	100	[0.0056, 0.0117]	0.0110	0.1701
1000	10	[0.0050, 0.0109]	0.0084	0.1268
1000	25	[0.0043, 0.0081]	0.0083	0.1627
1000	100	[0.0046, 0.0100]	0.0085	0.1710

$r = |z|$, and $R = 0.25$. On region i define

$$u_i^\star = \mathrm{Re}\big[(z/R)^j\big] \exp[-a_i(r^2 - R^2)], \qquad a_i = \frac{j + 1/\kappa_i}{2R^2}, \qquad f_i = \kappa_i[4a_i(j+1) - 4a_i^2 r^2]u_i^\star. \tag{12}$$

These smooth fields solve $-\kappa_i \Delta u_i = f_i$, share the trace $\cos(j\theta)$, and satisfy $\kappa_i \partial_r u_i^\star = -\cos(j\theta)/R$ at the interface, so their outward conormal fluxes balance even with a conductivity contrast. The disc has $\kappa_F = 1$. The cases are $j \in \{1, 4, 8\}$ with $\kappa_P = 1$, and $j = 4$ with $\kappa_P = 10$. Exterior Dirichlet values come from the manufactured field.

Reference, Robin weight, and stopping. The mesh has 128 interface nodes and nominal $h = 0.015$; P1 loads and error integrals use quadrature order six. Independent sparse Dirichlet solves check the exact discrete reference and its flux balance, and an exact Neumann solve on Ω_P reproduces the reference trace from the reference flux. The Robin weight minimizes the spectral radius of the exact exchange over 201 logarithmically spaced values in $[1, 10^5]$. This gives $\alpha \approx 105.9$ with exact rate 0.793 for unit conductivity and $\alpha \approx 4467$ with rate 0.095 for the contrast case. The choice uses the exact operators and is a benchmark control, not a deployable tuning rule. Every run starts from zero trace and flux, performs sparse FEM Robin solves, and stops after at least ten exchanges once $\|\lambda^{k+1} - \lambda^k\|_\Sigma / (1 + \|\lambda^{k+1}\|_\Sigma) < 10^{-9}$, or at 2000 exchanges. Neither the true nor the predicted solution enters this rule.

Fixed features. For each of five seeds, draw 1024 sine features with Gaussian frequency scale 16 and uniform phases, append a constant, and use nested subsets of 128, 512, and 1024 features. The 4096 interior and 512 exterior samples are fixed within a seed, and the 128 interface nodes are shared. The head is fit to the normalized PDE $-\Delta u = f_P/\kappa_P$, the interface Neumann condition $\kappa_P \partial_{n_P} u = -q$, and the exterior Dirichlet condition, with mean-square weights $1, 100, 100$, column equilibration, and SVD cutoff 10^{-10}. Refitting the head for each supplied flux gives a fixed affine learned solver. Forcing and exterior data change between cases; the feature bank and samples do not. Changing only the feature count isolates the effect of enlarging the feature space under the same sampled objective. It does not model feature learning in an MLP.

All 60 feature runs pass the update test, within 10 to 90 exchanges, and stop within 2.0×10^{-8} of the limit predicted by (7). All fixed maps M have spectral radius at most 0.90 at unit conductivity and 0.18 in the contrast case. At mode 8 the median global L^2 error is 3.53, 0.69, and 0.15 for 128, 512, and 1024 features. With 128 features, modes 4 and 8 and the contrast case all have median global errors above one: the fitted space cannot represent these interface fields, yet the exchange settles.

Tolerance-limited classical solves. We replace the learned solver by Jacobi-preconditioned conjugate gradients on the FEM system of Ω_P, with the interface Neumann data and exterior Dirichlet data imposed exactly. Six relative residual tolerances, $10^{-1}, 3 \times 10^{-2}, 10^{-2}, 10^{-3}, 10^{-5}, 10^{-8}$, are crossed with zero and previous-solution initial guesses on all four cases (48 runs). All 48 pass the same update test. Loose tolerances leave trace errors up to 1.7%, and the initial guess changes the error. With a warm start at a loose tolerance, CG often accepts the previous field after a few iterations, the trace barely changes, and the exchange stops after the minimum of ten exchanges. At tolerance 10^{-8} every trace error is at most 1.0×10^{-8}. Premature stopping and history dependence are therefore not unique to learned solvers. The relative criterion used here differs from the initial-residual-relative criterion analyzed by Birken (2015).

Table 7: **Equal-amplitude interface modes expose errors hidden by a smooth trace.** Medians over five seeds at termination. Trace error is relative to the discrete FEM–FEM reference; global relative L^2 and energy errors integrate both fields against the analytic solution. Energy uses $\int \kappa|\nabla e|^2$. "Stops" counts runs that pass the update test.

j	κ_P	Features	Trace error	Global L^2	Energy	Stops
1	1	128	0.0109	0.0463	0.0948	5/5
1	1	512	0.0007	0.0017	0.0383	5/5
1	1	1024	0.0007	0.0017	0.0383	5/5
4	1	128	1.3478	5.1402	2.4758	5/5
4	1	512	0.0046	0.0096	0.0873	5/5
4	1	1024	0.0047	0.0094	0.0873	5/5
8	1	128	1.2342	3.5246	1.3809	5/5
8	1	512	0.1119	0.6902	0.8170	5/5
8	1	1024	0.0165	0.1537	0.3050	5/5
4	10	128	0.5237	1.4801	1.2036	5/5
4	10	512	0.0056	0.0083	0.0399	5/5
4	10	1024	0.0056	0.0083	0.0399	5/5

Table 8: **Inner tolerance and initialization matter for classical solves too.** Selected tolerances from the six-value grid. Both error columns are trace errors at termination against the same discrete reference; the last column is the global L^2 error of FEM on both regions.

j	κ_P	CG tolerance	Zero-start error	Warm error	Exact FEM L^2
1	1	1e-01	9.67e-03	1.20e-02	0.0021
1	1	1e-02	9.23e-04	8.30e-04	0.0021
1	1	1e-03	5.14e-04	1.33e-04	0.0021
1	1	1e-08	1.40e-09	1.21e-09	0.0021
4	1	1e-01	3.29e-03	2.16e-03	0.0116
4	1	1e-02	1.01e-02	3.64e-04	0.0116
4	1	1e-03	1.76e-04	4.01e-05	0.0116
4	1	1e-08	9.12e-10	6.74e-10	0.0116
8	1	1e-01	9.34e-03	1.65e-02	0.0357
8	1	1e-02	5.74e-03	1.62e-03	0.0357
8	1	1e-03	1.36e-03	1.84e-04	0.0357
8	1	1e-08	1.02e-08	6.80e-09	0.0357
4	10	1e-01	9.30e-03	1.35e-02	0.0112
4	10	1e-02	1.66e-02	1.42e-03	0.0112
4	10	1e-03	3.40e-04	1.87e-04	0.0112
4	10	1e-08	1.52e-09	1.40e-09	0.0112

Discretization and field error. Table 9 refines the FEM mesh and polygonal interface jointly. For mode 8, the global L^2 error of FEM on both regions falls from 0.487 to 0.0357. On the finest grid the 1024-feature hybrid error is 0.15, so the learned error remains substantial. On easier cases a hybrid can have lower field error than P1 FEM on both regions, because the approximations on Ω_P differ; this is not a timing comparison.

An independent residual check. For each feature run we evaluate the PDE residual on 2048 new interior points, normalized by the forcing norm, and the exterior mismatch on 512 new boundary points, normalized by the RMS of the learned interface trace. The score is the larger of the two and uses no reference solution. We also record the mismatch between the learned flux and the imposed flux at 256 held-out angles. It is not used in the score: it compares against a piecewise-linear interpolant of the FEM flux, whose interpolation error is nearly the same for every run. Among the 60 settled runs, 31 have global L^2 error above 1%. A score threshold of 1% flags 30 of them and one of the 29 others; the missed run has error 1.00%. This is a post-hoc illustration on one benchmark, not a certified estimator or a validated threshold.

Artifacts. The bundle contains 20 JSON records: twelve reference cases in one file, 60 feature runs in fifteen files, and 48 inexact-FEM runs in four files. Hashes identify the records and the executed

Table 9: Errors of FEM on both regions against the analytic fields under joint mesh and interface refinement.

j	κ_P	n_Σ	h	Trace error	Global L^2	Energy
1	1	32	0.060	0.0113	0.0366	0.1971
1	1	64	0.030	0.0025	0.0084	0.0943
1	1	128	0.015	0.0007	0.0021	0.0471
4	1	32	0.060	0.0622	0.1858	0.4140
4	1	64	0.030	0.0168	0.0457	0.2081
4	1	128	0.015	0.0047	0.0116	0.1049
8	1	32	0.060	0.1470	0.4865	0.6710
8	1	64	0.030	0.0488	0.1291	0.3420
8	1	128	0.015	0.0159	0.0357	0.1799
4	10	32	0.060	0.0479	0.1693	0.3402
4	10	64	0.030	0.0199	0.0441	0.1837
4	10	128	0.015	0.0056	0.0112	0.0919

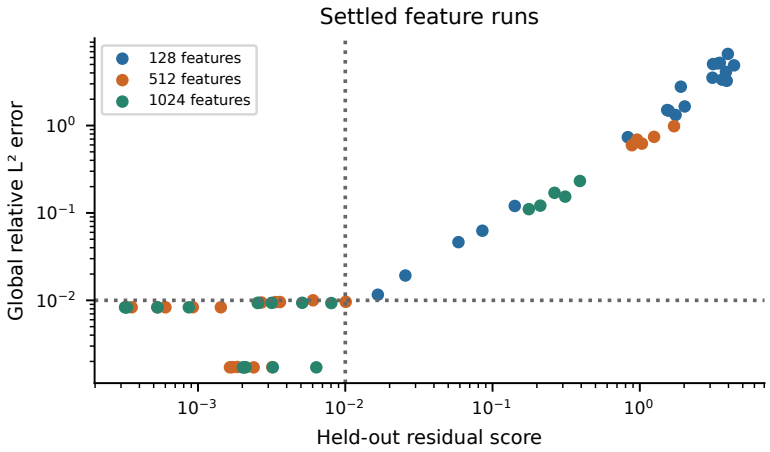

Figure 7: Held-out residuals separate inaccurate from accurate feature runs after both pass the same stopping test. Dotted lines mark the illustrative 1% thresholds.

source. The bundled analysis code verifies the complete grids and regenerates the summaries, figures, and tables, and the runnable driver with its dependencies is included. Smoke results are excluded.

F Experiment protocols and statistical conventions

F.1 Shared geometry and implementation

The square is $(0,1)^2$ and the disc has center $(0.5, 0.5)$ and radius $R = 0.25$. The base problem uses $u^\star = \sin(\pi x)\sin(\pi y)$, zero exterior Dirichlet data, and $f = 2\pi^2 u^\star$. Subdomain FEM uses P1 triangles, nominal size $h = 0.06$, and 32 matched interface vertices; the meshes have 63 disc vertices with 92 triangles and 279 outer vertices with 458 triangles. Interior collocation points are sampled in the relevant region, and PDE derivatives are obtained by automatic differentiation. PINNs use a width-64, depth-5 tanh architecture in float32; random-feature linear systems use float64. The software environment is Python 3.14, NumPy 2.4, SciPy 1.17, and PyTorch 2.11; figures are generated with Matplotlib.

F.2 Disc response estimators

All estimators train a PINN with 1500 Adam steps, learning rate 10^{-3}, 512 interior points, and interface penalty weight 100 per probe. Fourier traces probe modes through $j = 4$; the comparison

reports $j = 1, 2, 3$. The nodal and random estimators build a 32-dimensional response matrix, whereas the Fourier method targets modal projections. Each estimator and mode has five seeds. We report the median and maximum relative error.

F.3 Robin-weight sweeps

Rate sweeps use ten exchanges and 1500 Adam steps per exchange. The disc-PINN sweeps use 512 collocation points; the outer-region PINN sweep uses 1024. Each configuration starts near the manufactured solution, with trace perturbation $0.5\mathcal{N}(0, I)$ from NumPy seed $42 + s$, network seed s, and exact manufactured initial flux. The disc protocol pretrains for 600 steps and the outer-region protocol for 1500. Networks are warm-started across exchanges. Values are grouped at four significant digits and checked for one observation per seed per setting.

The reported rate is the geometric mean of consecutive manufactured-trace error ratios; values at or below 10^{-10} are removed, and ratios outside $(10^{-6}, 10^{6})$ are discarded. This finite-window metric includes training transients, discretization error, and neural bias. A rate above one indicates growth of this metric. Spreads are population standard deviations.

F.4 FEM–FEM sweep with arbitrary initialization

The disc receives the Robin data and the outer region the Neumann data. Both Schur matrices are assembled on the matched interface with a consistent interface mass. The exchange runs forty steps at each of fifteen values of α, starting from a trace perturbed with seed 42 and the manufactured flux. The tail rate uses ratios among the last eleven successive-update norms, omitting denominators below 10^{-13}. It measures the asymptotic factor once the dominant mode has emerged, and complements the dominant-eigenvector test of Appendix D.

F.5 Schwarz and stopping metrics

The overlap-0.05 configuration uses radii $r_1 = 0.225$ for the FEM hole and $r_2 = 0.275$ for the PINN disc; for overlap 0.10 these are 0.20 and 0.30. The PINN receives the FEM Dirichlet trace at r_2, and FEM receives the PINN trace at r_1. All methods use 1024 collocation points, 1500 Adam steps, learning rate 10^{-3}, and boundary weight 100. Each network starts fresh and keeps its weights between exchanges; the Robin–Neumann exchange starts with zero flux. Initial trace perturbations have amplitude 0.3. This transplants the algorithm of Snyder et al. (2023) onto our geometry; it does not reproduce their experiments.

For each seed, the first exchange at which the relative trace error is below the threshold is the first-hit time. Table 13 reports its median over successful seeds with their count; unsuccessful runs are not assigned a fictitious value. Minimum and step-25 errors are medians over all twelve seeds. Because the methods use different interfaces and overlap volumes, these are controlled trace comparisons, not matched global-error or equal-wall-time benchmarks.

F.6 Restart and budget study

The disc-PINN study uses $\alpha = 10.313$, sixteen exchanges, 512 interior points, Neumann penalty weight 100, and the shared width, depth, and learning rate. The warm budget grid is $\{10, 25, 50, 100, 250, 500, 1000, 2000, 4000\}$ and the cold grid $\{250, 1000, 4000\}$, each with seeds 0–11. Warm runs keep the network weights and construct a fresh Adam optimizer at each exchange; cold runs also reset the network. The initial trace is the manufactured trace plus $0.3\mathcal{N}(0, I)$ drawn with seed $1000 + s$, and the initial flux is the manufactured flux; neither arm uses pretraining. Last-six-exchange medians are taken separately for relative trace error and training loss, and the square root is applied to the latter. Fits use ordinary least squares of $\log(\text{error})$ on $\log(\sqrt{L})$ with intercept, and R^2 is computed in log coordinates. The conditional comparison uses the three common budgets and removes their mean effects in log coordinates. Repeated seeds across budgets are not independent replicates; Appendix H resamples whole seed clusters.

Table 10: **Fourier probes recover low-frequency interface flux accurately.** Relative error against the exact disc response j/R: median (worst) over five seeds. Counts include each protocol's offset handling.

Estimator	Trainings	$j = 1$	$j = 2$	$j = 3$
Fourier	5	5.3e−5 (2.4e−4)	1.3e−3 (2.4e−3)	1.1e−2 (1.8e−2)
Nodal basis	33	1.3e−2 (3.1e−2)	4.7e−2 (5.3e−2)	7.1e−1 (7.2e−1)
Random, 32 traces	33	1.0e−1 (1.7)	9.7e−2 (2.3)	2.8e−1 (9.6)
Random, 64 traces	65	7.0e−3 (1.1e−2)	9.9e−3 (1.4e−2)	7.7e−2 (1.1e−1)

G Validation of boundary-frequency probes

The probes of Section 4.3 target selected Dirichlet-to-Neumann responses. On the disc the exact value is $s_j = j/R$. A homogeneous problem with negligible zero-input response can omit the offset solve in (8).

Smooth probes isolate one boundary frequency. A nodal impulse excites many frequencies, while random traces support reconstruction of the full nodal map. Table 10 compares these protocols with 1500 Adam steps per boundary problem. Fourier probing gives smaller median errors on the first three modes with five trainings, versus 33 or 65. The counts reflect different targets—selected modal responses versus a full nodal map—and do not establish equal-information efficiency.

H Conditioning the loss–error relationship on budget

Let x_{sb} and y_{sb} be the log square-root loss and log terminal trace error for seed s and budget b, using the last-six-step summaries in Section 4.4. The partial R^2 is the squared correlation of $x_{sb} - \overline{x}_b$ and $y_{sb} - \overline{y}_b$, computed separately for each restart policy. Equivalently, it is the fraction of residual error variance explained by adding log loss to a regression containing categorical budget effects. The comparison uses the shared budgets 250, 1000, and 4000: twelve seeds at each budget for each policy. It is a descriptive association, not an out-of-sample prediction score.

Table 11: **A weak pooled fit need not mean weak within-budget association.** Brackets give percentile 95% intervals from 5000 seed-cluster bootstrap samples. Each draw resamples twelve seed identifiers with replacement and preserves all their budgets and both restart policies.

Policy	Pooled R^2	Partial R^2 given budget
Warm starts	0.875 [0.783, 0.931]	0.870 [0.758, 0.934]
Cold restarts	0.198 [0.096, 0.306]	0.760 [0.611, 0.843]

The paired bootstrap interval for warm-minus-cold partial R^2 is $[0.018, 0.202]$, with point estimate 0.111. Warm starts retain the stronger association in this comparison, but the cold association is substantial. Separate cold fits at budgets 250, 1000, and 4000 have slopes $3.045, 2.201, 1.518$ and $R^2 = 0.478, 0.822, 0.873$, respectively. The full warm grid (108 runs over nine budgets) has pooled slope 1.110 and $R^2 = 0.884$; the shared-budget warm grid has slope 1.537 and $R^2 = 0.875$. These changes caution against interpreting the pooled slope as a universal scaling exponent.

The study initializes traces with perturbed manufactured values and starts from manufactured flux (Appendix F); it is not an evaluation of deployment from unknown interface data. A prediction claim would require evaluation on held-out seeds or budgets. The executable analysis verifies every terminal summary against its sixteen recorded iterations, and the bundled records contain all fits and bootstrap settings.

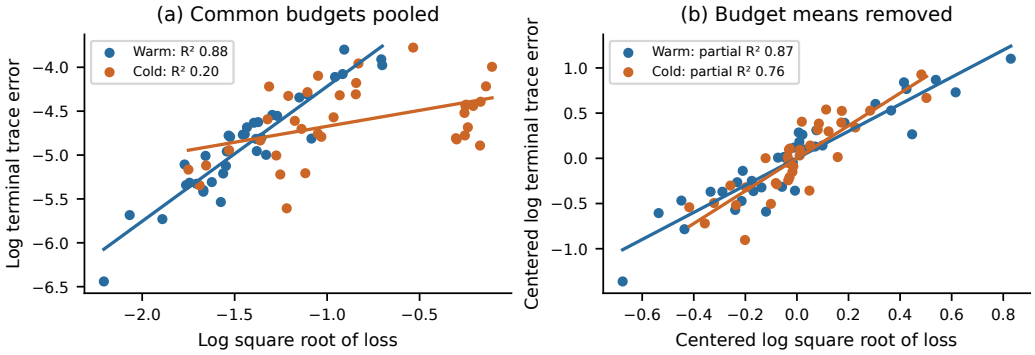

Figure 8: The same 36 runs per policy before and after removing each budget's mean in log coordinates. Conditioning changes the interpretation of the cold-restart association.

I FEATURE CONTROLS FOR INTERFACE-RESPONSE FIDELITY

For a circular annulus with inner radius R, outer radius R_o, and zero exterior trace, the exact response to angular mode $j \geq 1$ is

$$s_j^{\text{circ}} = \frac{j}{R} \frac{R_o^{2j} + R^{2j}}{R_o^{2j} - R^{2j}}. \tag{13}$$

Our region has a square outer boundary, so the reference instead solves homogeneous Dirichlet FEM problems on the actual region at $(n_\Sigma, h) = (64, 0.03), (128, 0.015), (256, 0.0075), (512, 0.00375)$ and projects each returned outward flux onto the imposed mode. The finest responses for $j = 1, 2, 3$ are approximately 6.19075, 8.94136, and 12.25090. The last refinement changes them by at most 0.21%, and the responses through $j = 8$ by at most 1.24%. These are finite-resolution modal projections, not eigenvalues of a simultaneously diagonalized operator. Neural signs are converted to the region's own outward convention before signed relative errors are computed; attenuation uses magnitudes. The low-band random-feature LS discrepancy of 0.15% is near the reference's refinement scale. The bundle also reports cap sensitivity at thresholds 0.7, 0.8, and 0.9.

The probes train each prescribed mode independently and project the total returned flux. Because the problems are homogeneous, no separate zero-input baseline is trained, so any neural offset is included in the measured response. They are empirical boundary-response diagnostics rather than measurements of a fixed affine map.

The response probe uses homogeneous Laplace forcing, the prescribed interface Fourier trace, and zero Dirichlet data on the square boundary. Random features are sinusoids with randomly drawn frequencies; the main comparison uses 1600 features with frequency scale 16. For both random-feature fitting methods the collocation system has 1536 interior points, 32 interface points, and 512 exterior points. The PDE block has multiplier 1 and both boundary blocks multiplier 100 *before* squaring the stacked least-squares residual. Columns are equilibrated by their Euclidean norms. LS uses NumPy's least-squares solver with relative rank threshold 10^{-10}; Adam fitting of the same equilibrated system uses learning rate 10^{-2} and 2500 steps.

The learned-feature baseline trains the whole tanh MLP with 1024 interior collocation points, boundary weight 100, learning rate 10^{-3}, and 2500 steps. Its refit variant repeats that training, freezes the penultimate features, and solves an additional collocation least-squares problem in float64; the fitted coefficients are written back into the float32 network. The refit changes the collocation objective and adds work, so it tests whether the frozen feature span supports a better fit rather than providing a balanced timing comparison. Its residual near 0.98 at high modes is specific to the sampled system and rank threshold.

All reported aggregates use exactly five seeds and 2500 Adam steps where applicable. Two single-seed runs with larger budgets (10000 steps for the learned network and 200000 for random-feature Adam) have cap indices of 5 and 4 against the circular reference, indicating that more optimization can move the threshold on fixed architectures; one seed does not establish a scaling law.

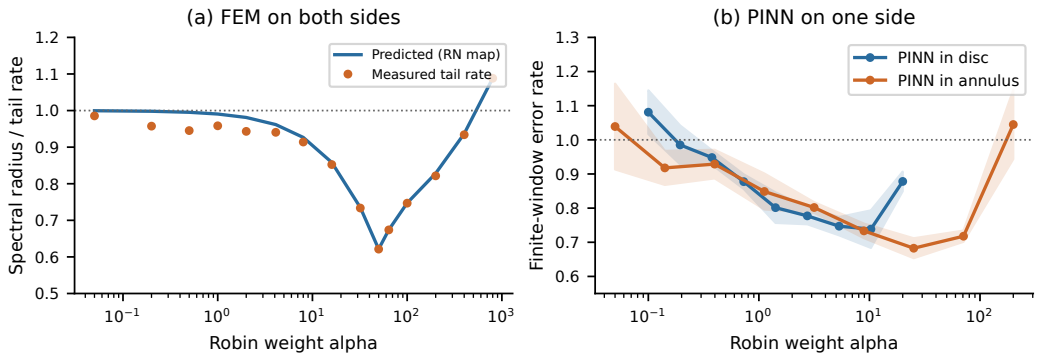

Figure 9: **Robin-weight sweeps.** (a) FEM on both regions: the spectral radius of (3) predicts the measured tail rate. (b) PINN on one region: finite-window error rates, mean ± standard deviation over five seeds. Values below one indicate average error reduction; they include retraining effects and are not fixed-operator rates.

A separate single-seed LS run with 3200 features, frequency scale 32, 4096 interior points, and 1024 exterior points recovers modes $j = 2, \ldots, 8$ within 0.1–2.8% of the circular-annulus curve, while $j = 0, 1$ differ by about 9.8% and 7.1%. The corresponding domain still has a square outer boundary. The LS systems are numerically rank deficient (ranks about 1291–1333 out of 3200), so this run shows attainable sampled responses, not global PDE accuracy. Increasing both feature count and scale does not isolate either variable.

J Robin-weight sweeps with PINNs

FEM on both regions. Figure 9(a) compares the spectral radius of (3) with forty-step exchanges started from a perturbed trace. At $\alpha = 50$ the predicted rate is 0.621426 and the measured tail rate 0.621300; at $\alpha = 800$ both are approximately 1.08834, indicating growth. The relative discrepancy is below 1% at the eight sampled weights from 16 to 800. Slower runs at small α need a longer window to reach the asymptotic rate (Table 12).

Table 12: **Predicted and tail rates from arbitrary initialization**, six of fifteen weights. Relative discrepancy is $|r_{\text{tail}} - \rho(T)|/\rho(T)$; small-$\alpha$ discrepancies reflect the finite window.

α	$\rho(T)$	Measured tail rate	Relative discrepancy (%)
0.05	0.999524	0.985462	1.407
2	0.981143	0.943136	3.874
16	0.858297	0.852420	0.685
50	0.621426	0.621300	0.020
200	0.828375	0.821594	0.819
800	1.088349	1.088341	0.001

PINN on one region. With retraining, we report the geometric mean of consecutive trace-error ratios over the recorded window (Figure 9b). With the PINN on the disc, the Dirichlet–Neumann exchange gives 1.477 ± 0.292 and the Robin–Neumann exchange at $\alpha = 4.16$ gives 0.745 ± 0.021. With the PINN on the outer region, Dirichlet–Neumann gives 1.439 ± 0.149 and Robin–Neumann 0.683 ± 0.029 at $\alpha \approx 25.1$. In the latter configuration every sampled α from about 0.141 to 70.9 gives a rate below one on every seed; $\alpha = 200$ does so on only one of five. A twelve-configuration transport study gives Dirichlet–Neumann rates of 1.21–1.82 and Robin–Neumann rates of 0.73–0.77 at $\alpha = 10$ (Appendix M). These sweeps start near the known solution and support the practical value of Robin transmission without identifying an optimal weight.

Table 13: **Reaching a tolerance first and finishing with the smallest error give different rankings.** Twelve seeds. First-hit counts are medians over successful seeds, with the success count in parentheses; minimum and final errors are medians over all seeds. Errors use each scheme's own interface. †The Dirichlet–Neumann exchange is stopped by the driver's growth criterion after 4–6 exchanges; its entry is the last recorded error.

Method	Exchanges to 10^{-2} (successes)	Minimum error	Error at exchange 25
Dirichlet–Neumann	— (0/12)	6.1×10^{-1}	$7.6^{\dagger}$
Schwarz, overlap 0.05	6 (7/12)	8.9×10^{-3}	2.1×10^{-2}
Schwarz, overlap 0.10	5.5 (12/12)	2.2×10^{-3}	1.0×10^{-2}
Robin–Neumann, $\alpha = 4.16$	17 (12/12)	1.6×10^{-3}	1.9×10^{-3}
Robin–Neumann, $\alpha = 10.31$	23.5 (6/12)	1.0×10^{-2}	1.0×10^{-2}

K Stopping metrics in a Schwarz comparison

Overlapping Schwarz exchanges Dirichlet values across an overlap, providing a comparison with the value–flux exchange of the nonoverlapping Robin–Neumann method. We adapt the weak-Dirichlet PINN coupling of Snyder et al. (2023) to our geometry with overlap widths 0.05 and 0.10, and include the unrelaxed Dirichlet–Neumann exchange as a baseline. All schemes start with a fresh network, use 1500 Adam steps and 1024 interior points per exchange, and run up to 25 exchanges over twelve seeds, without pretraining. We report the number of exchanges needed to first reach 10^{-2} relative trace error and the error after 25 exchanges.

Larger-overlap Schwarz reaches 10^{-2} sooner, but its final error exceeds its best error on every seed: stopping at a tolerance and stopping at a budget rank the methods differently. Robin–Neumann at $\alpha = 4.16$ needs about three times as many exchanges as large-overlap Schwarz to reach the tolerance, and has the lowest median final error. A separate relaxation comparison favors Aitken-accelerated Dirichlet–Neumann over the Robin–Neumann settings tested there (Appendix L). We therefore make no claim that Robin transmission is uniformly best.

L Interface-relaxation baselines

A separate disc-in-square comparison evaluates unrelaxed Dirichlet–Neumann, a tuned constant relaxation, scalar Aitken relaxation, IQN-ILS, and two Robin–Neumann settings over five seeds. IQN-ILS applies a history-based interface quasi-Newton update to the Dirichlet–Neumann map with startup relaxation 0.1 and least-squares rank threshold 10^{-8}. These runs use their own initialization and a twelve-exchange protocol, so their absolute errors must not be combined with the Schwarz comparison. The constant relaxation and the searched Robin weight are selected on seed 0 before evaluation across seeds.

Table 14: Relaxation comparison, mean $\pm$ standard deviation over five seeds. No method reaches the 0.05 relative trace-error tolerance within the budget. The Robin weights are tested settings, not validated optima.

Method	Final relative trace error	Finite-window rate
Dirichlet–Neumann	$(3.6 \pm 3.5) \times 10^{3}$	1.50 ± 0.19
Dirichlet–Neumann, constant $\omega = 0.20$	0.27 ± 0.025	0.61 ± 0.0066
Dirichlet–Neumann, Aitken	0.20 ± 0.043	0.57 ± 0.014
Dirichlet–Neumann, IQN-ILS	4.2 ± 4.0	0.73 ± 0.074
Robin–Neumann, $\alpha = 1.44$	1.1 ± 0.26	0.84 ± 0.018
Robin–Neumann, $\alpha = 3.07$	0.76 ± 0.24	0.76 ± 0.032

Aitken outperforms the tested Robin–Neumann settings on this protocol. IQN-ILS has a favorable average rate but seed-sensitive final errors from about 0.22 to 11.4, another reason to report physical error together with a contraction statistic. The data do not show that quasi-Newton acceleration is generally unsuitable for PINN coupling.

M DIFFUSION-DOMINATED TRANSPORT

Replacing the Poisson operator by $-\varepsilon\Delta u + \mathbf{b}\cdot\nabla u$, with $\varepsilon = 1$, constant $\mathbf{b}$, the same manufactured solution, and adjusted forcing, tests the rate separation beyond symmetric problems. We use $n_\Sigma \in \{24, 32, 48\}$, nominal $h = 0.06$, and four velocities from $(0.5, 0.25)$ to $(4, 2)$. The mesh Péclet number $\|\mathbf{b}\|h/(2\varepsilon)$ ranges from about 0.017 to 0.134, so the problems are diffusion dominated. The Robin–Neumann exchange uses $\alpha = 10$, selected from the grid $\{0.3, 0.6, 1.2, 2.5, 5, 10, 20, 40\}$, with three seeds per configuration. This is a modest extension of the observed rate separation, not a validated nonsymmetric theory.

Table 15: Finite-window rates for the transport extension, mean $\pm$ standard deviation over three seeds.

n_Σ	Mesh Pe	Dirichlet–Neumann	Robin–Neumann ($\alpha = 10$)
24	0.017	1.77 ± 0.31	0.74 ± 0.08
24	0.034	1.78 ± 0.18	0.74 ± 0.07
24	0.067	1.75 ± 0.33	0.73 ± 0.07
24	0.134	1.82 ± 0.32	0.73 ± 0.06
32	0.017	1.37 ± 0.14	0.75 ± 0.03
32	0.034	1.51 ± 0.15	0.77 ± 0.03
32	0.067	1.42 ± 0.10	0.75 ± 0.03
32	0.134	1.56 ± 0.21	0.75 ± 0.04
48	0.017	1.53 ± 0.25	0.75 ± 0.06
48	0.034	1.21 ± 0.05	0.77 ± 0.03
48	0.067	1.35 ± 0.11	0.76 ± 0.03
48	0.134	1.48 ± 0.25	0.74 ± 0.06

N ARTIFACT MAP AND NUMERICAL VERIFICATION

The source bundle contains the manuscript source, the bibliography, the unmodified ICLR style files, vector figures, the experiment records, and the analysis code. The analysis reads only the bundled records, checks seed counts and record hashes, reproduces all aggregates, and draws every figure, including Figure 1. No neural training is required to rebuild the figures.

Table 16: Result-to-experiment mapping. Experiment IDs identify the corresponding solver drivers in the source bundle.

Result	Experiment
Robin–Neumann verification	E33
Controlled modes, inexact FEM	E34
Disc probe	Estimator study
Feature controls	E29
Square-region response	E32
FEM–FEM sweep	E9
Robin-weight sweeps	E5/E9/E10
Schwarz comparison	E26
Loss/restart study	E25
Relaxation baselines	Relaxation, IQN-ILS
Transport extension	E28

Thirty random noncommuting-matrix checks verify the recursion (5), the α-independent limit (7) and its Dirichlet-to-Neumann form, and the measured-flux variant of Appendix C, each at three Robin weights. Automated tests check the exact exchange map against PDE sweeps and the limit of a fixed learned solver at two weights.