Neural Boundary Integral Operators: Solver-Consistent Learning Across Geometry, Resolution, and Kernel Mismatch*

Anonymous authors
Paper under double-blind review

Abstract

Neural operators typically learn PDE solution maps without preserving the singular kernel structure or discretization properties of the underlying numerical problem. We introduce the Neural Boundary Integral Operator (NBIO), a differentiable boundary-integral layer that retains the analytic singular kernel, solves a second-kind boundary integral equation for the boundary density, and evaluates the resulting potential at arbitrary query points. With zero learned correction, NBIO reduces to the classical Nyström method: on 2D Laplace problems it reaches relative L_2 error 4.8×10^{-7} and continues to converge under boundary refinement, while same-data neural surrogates remain near 10^{-2}–10^{-1}. To test a regime in which learning is essential, we withhold the modified-Helmholtz screening law and learn a scalar multiplier $m_\theta(\kappa\|x-y\|)$ of the retained Laplace normal form. In matched-capacity experiments, the radial model reaches 1.51×10^{-4} on unseen parameters, geometries, and boundary loads, compared with 1.18×10^{-2} for a generic pairwise correction; an independent experiment reproduces the result. The same parameterization reaches 1.07×10^{-4} when the entire interval $\kappa \in (1.5, 2.5)$ is withheld from training, compared with 1.91×10^{-3} for the stronger radial control, and outperforms a same-protocol Boundary Transformer. It also preserves the expected scale covariance, supports differentiable inverse recovery, and combines with split QBX for near-boundary evaluation. NBIO therefore separates two roles of structure: the boundary-integral core provides numerical consistency, while symmetry- and dimension-aware learning identifies missing kernel physics within a prescribed family. The current results address interpolation within screened kernel families rather than arbitrary kernel discovery.

1 Introduction

Operator learning aims to approximate solution maps of PDEs, $\mathcal{G} : a \mapsto u$, from data, enabling fast surrogate solves across a family of inputs. Fourier Neural Operators (Li et al., 2021) and DeepONet (Lu et al., 2021) are the dominant paradigms, with a now-substantial theory of neural operators between function spaces (Kovachki et al., 2023). These are powerful but *black-box*: they learn a volumetric map on a (near-)regular grid, do not use the structure of the differential operator, staircase on curved boundaries, and represent unbounded exterior domains only through truncation, mapping, or a separate exterior treatment. Variants address geometry through graph kernels (Li et al., 2020), learned deformations (Li et al., 2023a), or 3D geometry encodings (Li et al., 2023b), but the volumetric, structure-agnostic nature remains.

For *linear elliptic* PDEs, classical numerical analysis offers a powerful alternative: the *boundary integral equation* (BIE) reformulation (Atkinson, 1997; Kress, 2014; Hsiao & Wendland, 2008). Using the Green's function, the interior solution is represented by an integral of an unknown density over the $(d-1)$-dimensional boundary; the density solves a well-conditioned second-kind Fredholm equation. BIE methods are dimension-

*Large language models were used as assistive tools for manuscript editing, code review, and experiment-planning support. The authors designed and verified all methods, experiments, results, and claims, and take full responsibility for the content.

reduced, spectrally accurate (Nyström (Nyström, 1930) with Kress quadrature (Kress, 1991)), and exact for exterior problems. This structure is absent from black-box operators.

Contributions and questions. NBIO is a hybrid operator architecture that reduces to a classical boundary-integral solver when the governing kernel is known and uses learning for nonsingular model discrepancy when the kernel is misspecified or unknown. It assembles the known singular boundary operator directly and restricts data-driven capacity to a scalar prior-strength gate and a nonsingular pairwise correction, or, for the screened family, to a dimensionless radial multiplier of the retained source-normal kernel. We organize the evidence around three questions.

Q1: What is obtained from exact structure? The structure-only limit matches the 2D Laplace solver, tracks quadrature refinement and geometry shift, and extends to exterior, transmission, interface, and CAD problems. Its 10^2–10^5 error ratios over same-protocol surrogates are comparisons between a per-instance structured solve and amortized learned predictors, not evidence that a trained correction learned a better Laplace operator. *Q2: When does learning add value?* A fixed-parameter wrong-kernel test first shows that an unconstrained correction can remove discrepancy but may be hindered by a globally wrong prior. Our main learned experiment instead retains only the correct singular normal form and learns a dimensionless radial screening law over a parameter family. Parameter-matched controls isolate radiality from the κr collapse, and frozen tests cover geometry, load, resolution, query-distance, and scale shifts. A variable-coefficient study remains an explicitly related volumetric control. *Q3: Can the layer be used as a numerical component?* We test gradients in a shape inverse problem, repeated-right-hand-side reuse, close evaluation, and fast structured linear algebra; broader 3D, nonlinear, and transient examples are deferred to the appendix. Prop. 2 gives a fixed-geometry sufficient stability result; it does not assert that an arbitrary checkpoint satisfies the required margin.

Throughout, *zero-shot* means no update to learned parameters. Geometry and resolution transfer still recompute geometry-dependent quadrature, assemble or apply a new operator, and solve for a new density. The method is therefore a differentiable structured solver layer rather than a fixed-cost feed-forward surrogate. Our claim "across geometry, resolution, and kernel mismatch" is consequently two-part: the BIE core supplies discretization consistency, while learned corrections are tested for compatibility with new geometries and discretizations at fixed κ and, when explicitly conditioned, at held-out κ inside the trained family. The radial normal-form experiment is an interpolation result over a prescribed screened-kernel family, not a claim of unrestricted PDE-parameter extrapolation or generic unknown-physics recovery.

2 Background and Related Work

Neural operators. FNO (Li et al., 2021), DeepONet (Lu et al., 2021), GNO (Li et al., 2020), Geo-FNO (Li et al., 2023a), GINO (Li et al., 2023b); theory in Kovachki et al. (2023); fair-comparison practice in Lu et al. (2022). PINNs (Raissi et al., 2019) solve single instances rather than the operator. **Learning Green's functions** is related but distinct: Boullé et al. (2022); Gin et al. (2021) learn global Green's functions, and BIN-G (Negi et al., 2024) learns a domain-independent radially symmetric Green function using an RBF network and PDE/BIE residual losses. The radial NBIO specialization does not claim the first learned radial Green function. It retains the analytic singular source-normal form, learns an endpoint-normalized dimensionless multiplier from solution data, and solves the density numerically for each new geometry and load. **Boundary integral methods**: second-kind theory (Atkinson, 1997; Kress, 2014; Hsiao & Wendland, 2008), logarithmic (Kress) quadrature (Kress, 1991), fast evaluation (Greengard & Rokhlin, 1987), and close/near-singular evaluation by Quadrature-by-Expansion (Klöckner et al., 2013), which we make differentiable and embed in the operator. Our strongest black-box baseline is an attention-based boundary operator in the spirit of Vaswani et al. (2017).

Boundary- and interface-aware neural operators. Several recent operators target boundaries or interfaces directly. IONet (Wu et al., 2024) and XI-DeepONet (Bi et al., 2025) use domain decomposition or level-set/interface conditioning to learn elliptic *interface* solution operators, reporting rel-$L_2 \sim 10^{-3}$–10^{-2} across 2D–6D with volume sources and interface jumps; the concurrent ϕ-DeepONet (Roy et al., 2026) uses a multi-branch DeepONet with nonlinear latent interface embedding, achieving rel-$L_2 \approx 2.7 \times 10^{-3}$ on 2D

interface problems (comparable to IONet, and two orders better than standard DeepONet). BENO (Wang et al., 2024) embeds boundary values and geometry into a graph/message-passing operator for elliptic PDEs with body forcing on irregular corner domains.

The closest concurrent work to NBIO is in the BIE-learning subspace. LOB (Fang et al., 2024) reformulates the PDE as a BIE and trains a physics-informed operator from boundary points only, reporting $\approx 6 \times 10^{-2}$ on 2D parametric Laplace. MAD-BNO (Wu et al., 2026a) further trains a boundary operator using only data synthesized from fundamental solutions (no solver required), reaching rel-$L_2 \approx 6 \times 10^{-3}$ on 2D Laplace; the BIE evaluation step recovers interior from boundary. BI-TDONet (Meng et al., 2024) uses SVD of the second-kind BIE operator to learn interior-to-boundary-to-interior maps, reporting rel-$L_2 \approx 4$–7×10^{-4} MRE on interior/exterior Dirichlet problems with a single training. BINO (Wu et al., 2026b) learns a geometry-aware Green's traction kernel for mesh deformation (different application domain: linear elasticity). Mousavi et al. (2026) study boundary-condition encodings for arbitrary backbones. Most directly related to our contraction and nonlinear-BIE claims, the concurrent FREDINO (Georgiou et al., 2026) learns *provably contractive* second-kind Fredholm integral operators and applies them to nonlinear elliptic PDEs via a BIE formulation; NBIO differs by keeping the singular kernel's *functional form* analytic (learning either a scalar prior-strength gate plus a nonsingular pairwise correction or a dimensionless radial multiplier) rather than learning the full integral operator, and by assembling and solving the discretized second-kind system.

Concurrent differentiable-BEM and derivative-operator work (2026). Differentiable boundary element methods and quadrature-aware boundary operators are an active 2026 direction, and we do not claim novelty for any of these ingredients in isolation. JAX-BEM (Hipperson et al., 2026) demonstrates a differentiable BEM for gradient-based acoustic shape optimisation; Shape-DINO (Gong et al., 2026) trains shape (Fréchet) derivatives alongside states and reports reliable optimisation gradients; and CLBO (Khan & Yao, 2026) builds a complex-linear, quadrature-aware surface operator for boundary-to-field acoustic prediction. Relative to these, NBIO's distinct contribution is a *conjunction*: an analytic singular-kernel split (initialising at the classical solver), a differentiable second-kind *density* solve, and continuous/QBX potential evaluation, assembled into one interface whose *solver consistency* is measured across geometry, discretization, and interior/exterior/interface regime – and whose behaviour is tested under kernel mismatch and changing geometry. We claim neither differentiable BEM, nor quadrature-aware boundary operators, nor boundary-based electrostatics as novel on their own.

NBIO's novelty is a *conjunction*, not assembled differentiable BEM on its own. Relative to learned boundary operators (LOB, MAD-BNO, BI-TDONet), which train around a BIE but do not assemble and solve the geometry-dependent density system, NBIO explicitly assembles and solves it at inference – inheriting the classical solver's reference-level accuracy (rel-$L_2 \sim 10^{-7}$ on 2D Laplace, vs 10^{-3}–10^{-4} for the closest BIE-learning competitors, reported under different protocols) and zero-shot resolution transfer. Relative to classical *differentiable* BEM (e.g. JAX-BEM), which already assembles, solves, and differentiates the exact-kernel system, NBIO adds a learned discrepancy parameterization (a nonsingular pairwise correction or the screened radial multiplier) and a shared continuous/QBX operator interface across geometries and discretizations; the reference-level accuracy itself is a property NBIO shares with such solvers, not a point of difference. On interface problems with jumps and volume sources, NBIO further generalizes the Calderón formulation and approaches reference-solver accuracy across the full input class; ϕ-DeepONet, IONet, and XI-DeepONet report far larger errors on related interface classes, but on different datasets and protocols (contextual reference points, not head-to-head). The trade-off is scope: the advantage is in the BIE-native regime; on volumetric variable-coefficient problems its edge over black-box FNO is modest.

3 Method: NBIO

We first define the boundary-integral architecture for a homogeneous elliptic Dirichlet problem. The variable-principal-coefficient experiment in §4.2 is a related volumetric structured-prior control, not an instance of this boundary-only architecture. Let $\Gamma = \partial\Omega$ be a positively oriented, sufficiently regular closed curve with

method	singular kernel	density solve	learned object	off-boundary query	resolution test
LOB (Fang et al., 2024)	BIE in formulation	no assembled solve	boundary map	BIE evaluation	not central
BI-TDONet (Meng et al., 2024)	BIE/SVD preprocessing	learned factor maps	solution maps	BIE evaluation	not reported here
JAX-BEM (Hipperson et al., 2026)	analytic	explicit	none	continuous	classical discretization
FREDINO (Georgiou et al., 2026)	learned operator	contractive iteration	full Fredholm map	formulation-dependent	not central
CLBO (Khan & Yao, 2026)	quadrature-aware	no density solve	boundary-to-field map	continuous field	surface discretizations
NBIO	**analytic**	**explicit**	**nonsingular corr. or radial law**	**continuous/QBX**	**measured**

Table 1: Method-level positioning based on the reported formulations. "No assembled solve" does not mean that a method ignores BIE physics; it distinguishes learning a solution/boundary map from explicitly assembling and solving a geometry-dependent density equation at inference. JAX-BEM is a classical differentiable solver rather than a neural operator.

outward normal n_y. In 2D we use $G_0(x,y) = -(2\pi)^{-1}\log|x-y|$ and

$$(D_0\mu)(x) = \int_\Gamma \partial_{n_y} G_0(x,y)\mu(y)\, ds_y.$$

With this convention the interior trace is $(-\frac{1}{2}I + D_0)\mu = g$ and $u(x) = D_0\mu(x)$ for $x \in \Omega$. The exterior trace is $(+\frac{1}{2}I + D_0)\mu = g$; in 2D we add the standard rank-one deflation used by our implementation to remove the constant nullspace. Appendix C gives the discretization, exterior form, transmission/source variants, and pseudocode.

M1: analytically-split kernel. The learned kernel is $K_\theta = (1+\gamma)K_0 + K_\theta^{\mathrm{corr}}$, where K_0 is the exact double-/single-layer kernel, including its analytic diagonal, and $\gamma = \tanh(\zeta)$ is a scalar gate initialized at $\zeta = 0$. The correction is a pairwise MLP. The original fixed-κ experiments use unit displacement, distance, e^{-r}, normal contractions, and the target normal. Because normalized displacement is not continuous at the diagonal, the regularity-controlled conditioned experiment instead uses raw displacement, squared distance, e^{-r^2}, and smooth normal contractions; this makes the composite correction smooth whenever the boundary parameterization is smooth. Both use nine geometric features (plus κ when conditioned), identical width/depth, and a near-zero final-layer initialization. The architecture contains a decaying feature, not a hard multiplicative envelope or spectral projection. Consequently boundedness is checked empirically and assumed, rather than enforced, in Prop. 2. Relative coordinates make the correction translation invariant; rotation invariance, reciprocity, kernel symmetry, and exact reparameterization invariance are not imposed.

Dimensionless radial normal form. When discrepancy belongs to an isotropic screened family, the generic pairwise correction discards known symmetry. We instead retain the Laplace double-layer normal form and learn a scalar multiplier,

$$K_{\theta,\kappa}(x,y) = K_0(x,y)m_\theta(s), \qquad s = \kappa|x-y|, \tag{1}$$

where $m_\theta(s) = 1 + s\, f_\theta(s, \log(1+s))$. This enforces $m_\theta(0) = 1$, leaving the leading local singular coefficient unchanged. It does not enforce $m'_\theta(0) = 0$: the three selected checkpoints have derivatives -0.0400, -0.0414, and -0.0515. The induced off-surface remainder is bounded but may have a direction-dependent diagonal limit; Appendix B reports two endpoint-matched variants that remove this defect but lose accuracy. We therefore do not call the learned remainder continuous across the diagonal. The exact modified-Helmholtz multiplier is $m_\star(s) = s\mathrm{K}_1(s)$, but this formula is withheld from training and used only for the numerical reference and post-hoc law audit. The three-layer width-32 MLP has 2,241 trainable parameters. A capacity-matched generic pairwise correction has 2,221 parameters; a second radial control has the same 2,241 parameters but receives $(\kappa/4, r)$ separately. Equation (1) is used in both matrix assembly and query evaluation. It is a family-specific inductive bias, not the default NBIO architecture for anisotropic discrepancy.

The capacity-matched generic controls use the same smooth pair features $\phi(x,y)$ in assembly and query evaluation:

$$K_{\mathrm{generic}}^{\mathrm{prior}} = K_0 + (\kappa/4)^2 C_\theta(\phi(x,y), \kappa/4), \qquad K_{\mathrm{generic}}^{\mathrm{none}} = C_\theta(\phi(x,y), \kappa/4).$$

The full-prior control retains the analytic Laplace diagonal and has no learned prior gate. The prior-free control learns the complete kernel, including its self-interaction, without a κ envelope. Both use their stated formula in the assembled density system and off-boundary potential.

Proposition 1 (scale covariance of the radial normal form)**.** *Let $x' = ax$, $y' = ay$, and $\kappa' = \kappa/a$ for $a > 0$. For the two-dimensional double layer, $K_0(x', y') = a^{-1}K_0(x, y)$, $ds_{y'} = a\, ds_y$, and $\kappa'|x' - y'| = \kappa|x - y|$. Hence the integral operator in (1) is unchanged under joint geometry/parameter scaling, and its potential satisfies $u'(ax) = u(x)$ for correspondingly scaled boundary data and query points.*

The proposition follows by substitution and makes no approximation claim. Section 4.2 tests whether the implementation preserves this covariance and whether the normal form improves statistical transfer relative to equally sized controls.

M2: differentiable second-kind solve. For nodes $y_j = x(t_j)$ and arc weights $w_j = |x'(t_j)|\Delta t$, we assemble $A_{ij} = \sigma\delta_{ij} + K_\theta(y_i, y_j)w_j$, with $\sigma = -1/2$ inside and $+1/2$ outside, and solve $A\mu = g$. The solve uses an automatic-differentiation-compatible dense solve or the damped second-kind fixed point $\mu_{k+1} = (1-\omega)\mu_k + \omega\{g - (A - \sigma I)\mu_k\}/\sigma$, with ω selected so that the tested discrete iteration matrices are contractive.

M3: continuous potential evaluation. At a query x_q we compute $u_h(x_q) = \sum_j K_\theta(x_q, y_j)w_j\mu_j$. The same pairwise function can be sampled on a different boundary discretization or at new query coordinates. This makes the *representation* independent of a fixed grid, but accuracy still requires a valid boundary parameterization, adequate quadrature resolution, interface separation, and a useful known kernel. No retraining is used in a resolution-transfer test, but a new matrix and density are computed.

Transmission, interface jumps, and volume source. For piecewise-constant media we use an indirect single-layer representation with Neumann–Poincaré coupling (contrast enters as known physics). For interface problems with prescribed jumps $[u] = \alpha$, $[a\partial_n u] = \beta$ we adopt the *direct (Calderón) formulation*: Cauchy data $(u, \partial_n u)$ are unknowns coupled across the interface through the single and double layers – no hypersingular operator is needed. The system is validated to machine precision ($\sim 10^{-15}$) against manufactured piecewise-harmonic solutions. For a volume source f, a particular solution $u_p = N[f]/a_k$ is computed via a non-singular FFT Poisson solve; the induced interface jump is absorbed into the homogeneous problem exactly, reducing the inhomogeneous case to the jump solver above. In 3D we use the Laplace single layer $1/(4\pi|x - y|)$ with a learned surface correction. For near-touching interfaces we evaluate the known layer potential with differentiable QBX (Klöckner et al., 2013) and oversample the near-singular cross-block assembly; the nonsingular learned correction stays a direct evaluation, preserving differentiability and continuous-query evaluation.

3.1 A fixed-geometry sufficient condition for stable correction

Let $X = C(\Gamma)$ with the supremum norm and let $A_0 : X \to X$ denote an invertible second-kind boundary operator on one fixed boundary Γ. Write $T_\theta : X \to X$ for the operator induced by the *complete* learned kernel perturbation – the scalar gate on the retained singular kernel together with the nonsingular learned correction, i.e. $\gamma K_0 + K_\theta^{\text{corr}}$ – and $A_\theta = A_0 + T_\theta$, so learning calibrates a scalar coefficient of the analytic singular kernel and adds a nonsingular correction term while the singular *functional form* is retained exactly and never itself learned. For the radial specialization, T_θ denotes the operator induced by $K_0(x, y)[m_\theta(\kappa|x - y|) - 1]$. Let $P_\theta : X \to Y$ be the corresponding potential evaluation operator into a normed query-space Y. The physical target operators are $A_\star = A_0 + T_\star$ and $P_\star$. The statement below separates stability of the perturbed equation, discretization error, and learned model discrepancy.

Proposition 2 (fixed-geometry stability and error decomposition)**.** *Assume $\eta = \|A_0^{-1}T_\theta\|_{X\to X} < 1$. Then A_θ is invertible and*

$$\|A_\theta^{-1}\|_{X\to X} \le \frac{\|A_0^{-1}\|_{X\to X}}{1 - \eta}.$$

Assume further that the integral parts $K_{\theta,h}$ form a collectively compact and consistent approximation. Then, for sufficiently small h, the full Nyström operators $\mathcal{A}_{\theta,h}$ and the corresponding collocation matrices $\mathbf{A}_{\theta,h}$ are

invertible with uniformly bounded inverses in the stated norms. If $\mu_\star$ solves $A_\star\mu_\star = g$ and $u_{\theta,h} = P_{\theta,h}\mu_{\theta,h}$, then

$$\|u_\star - u_{\theta,h}\|_Y \le C_{\text{stab}}\Big[\varepsilon_{A,h} + \varepsilon_{P,h} + \big(\|T_\theta - T_\star\|_{X\to X} + \|P_\theta - P_\star\|_{X\to Y}\big)\|\mu_\star\|_X\Big],$$

where the ε terms are absolute operator and potential quadrature errors (already evaluated on $\mu_\star$, hence not *rescaled by $\|\mu_\star\|$) and C_{stab} depends on the inverse margin and bounded potential norms. If the periodic parameterization, data, and both correction kernels have p derivatives, the quadrature terms have the corresponding algebraic order; for analytic periodic quantities they decay exponentially. Analytic or spectral convergence is therefore not implied by calling a neural correction merely "smooth." For a known-kernel target, $T_\star = 0$ and $P_\star = P_0$; setting $T_\theta = 0$ and $P_\theta = P_0$ then reduces the statement to convergence of the classical Nyström solver.*

A proof and the discrete-stability argument (collectively compact Nyström approximation) are in Appendix D. The first bound is the Banach perturbation lemma. Collectively compact approximation gives eventual discrete stability in the collocation norm (Atkinson, 1997); it does *not* by itself imply a uniform Euclidean matrix condition number for arbitrary node scalings. The error bound follows by adding and subtracting $P_\theta\mu_\star$, applying the resolvent identity to $A_\theta^{-1} - A_\star^{-1}$, and then adding the two quadrature errors.

This proposition is intentionally for one fixed Γ. Its constants may depend on the parameterization's minimum speed and regularity norms, curvature bounds, component separation, query distance, and the inverse stability of the continuum operator. A uniform statement over a geometry family requires uniform bounds on those quantities and on the correction/potential norms; the present proposition does not derive such a family-uniform theorem. The geometry-family tables below are therefore empirical transfer tests, not consequences of the fixed-geometry proposition.

The margin is a *sufficient* condition, not a guarantee that holds automatically for every trained checkpoint, and we separate three levels of evidence. *(a) Continuum theorem:* certification in the sense of Proposition 2 requires the $C(\Gamma) \to C(\Gamma)$ margin $\|A_0^{-1}T_\theta\| < 1$; the correction initializes near T_θ=0 (the exact classical solver), and a spectral-norm projection of the discrete correction would enforce the corresponding finite-dimensional condition if an a-priori guarantee were required (an available safeguard, not something these experiments run). *(b) Discrete certificate for the legacy generic correction:* we directly evaluate the discrete matrix margin $\|A_{0,h}^{-1}T_{\theta,h}\|_2$ using the *complete* learned perturbation $\gamma K_0 + K_\theta^{\text{corr}}$ (the scalar gate γK_0 contributes only ∼6% of it on average across all three seeds, at most 9%). Across all three trained seeds, four geometry families, and $n_b \in \{64, 128, 256\}$ (2304 audited matrices), it is 0.65 on average (median 0.65, p90 0.70) and at most 0.796, essentially independent of n_b; so the *finite-dimensional analogue* of the perturbation condition is satisfied on every audited matrix. This is stronger than condition-number evidence, but is not by itself a continuum $C(\Gamma)$-norm certificate (it would need an additional uniform norm-equivalence argument), and it is not guaranteed *a priori* – a stronger correction can push the margin past 1. *(c) Empirical transfer:* the assembled Euclidean condition number stays ≈ 2.5 across geometries and resolutions (§5.4), GMRES iteration counts are near-constant, and the error *decreases with n_b like a quadrature* (§4.1), supporting stability beyond the audited matrices; the black-box baselines show none of this refinement. A note on diagnostics: the earlier, looser product bound $\|A_{0,h}^{-1}\|_2\,\|T_{\theta,h}\|_2$ reaches 3.2 on these same cases, but that is only an upper bound on the margin; directly evaluating the tighter quantity $\|A_{0,h}^{-1}T_{\theta,h}\|_2$ gives the ≤ 0.796 above, so the product exceeding 1 does not contradict the measured margin.

The primary radial model has a separate certificate on an independently seeded holdout. Across five screening parameters and eight geometries per parameter for each of three checkpoints, the mean/p90/maximum $\|A_{0,h}^{-1}T_{\theta,h}\|_2$ are 0.406/0.464/0.499. The maximum condition-number ratio to the exact screened matrix is 1.0022, and the maximum relative solve residual is 2.58×10^{-7}. These are finite-dimensional sufficient-condition diagnostics, not a continuum family-uniform proof.

4 Experiments and Results

Ground truth is an exact manufactured solution where available, else a spectrally-accurate Nyström BEM (self-validated against analytic solutions to $4{\times}10^{-13}$). All compute on a shared GPU cluster. Because NBIO

is a hybrid architecture, the natural first question is *what is exact and what is learned* in each experiment; Table 2 is the reading guide.

experiment	exact core	learned component	train. params	interpretation
Interior Laplace (§4.1)	exact Laplace BIE	trained corr. + zeroed control	9k/0	structural validation
Fixed-κ wrong kernel	Laplace BIE (globally wrong)	generic correction	9k	fixed-mismatch baseline
Screened family (§4.2)	Laplace singular normal form	radial law $m_\theta(\kappa r)$	2.2k	**screened-law identification**
Variable coefficient	const.-coef. prior	residual operator	learned	volumetric scope test
CAD exterior (§5.1)	exact exterior BIE	tested, no benefit	0	known-kernel control
3D surface	approx. surface core	joint corr.	∼1.4k	geometry/regime transfer
Transmission/interface	exact multi-region BIE	none	0	multi-region structural validation
Numerical extensions	exact BIE/linear solve	none in headline	0	structure-only

Table 2: What is exact vs. learned in each experiment. Many headline results are the zero-parameter structural limit of the architecture (the exact classical solver). The main learned row withholds the screened constitutive law while retaining its singular normal form. Where the kernel is exact (interior Laplace, CAD, transmission/interface) a learned correction is unnecessary or, on CAD, mildly harmful (§5.1). *Naming convention used throughout:* a zero-parameter exact-core row is the *NBIO structural limit* (equivalently the classical BIE layer); a trained hybrid is *NBIO split+correction*; the no-split variant is the *learned-kernel BIE*. Solver-grade accuracy on exact-kernel rows is a property of the structural limit, not of learning.

4.1 Interior Laplace: reference-solver accuracy and zero-shot resolution

method	params	interior	n_b=64	n_b=128	n_b=256
classical BEM (reference)	—	4.4e−7	2.6e−4	7.8e−7	$<10^{-10}$
NBIO split+correction (trained)	**9k**	**4.79e−7**	2.6e−4	8.0e−7	9.5e−8
NBIO structure-only (zeroed corr.)	0 act.	4.75e−7	—	—	—
NBIO (Neumann solve)	9k	5.0e−7	—	—	1.3e−7
learned-kernel BIE (no split)	9k	2.3e−3	—	—	2.5e−3
Boundary Transformer	878k	2.4e−2	—	—	—
DeepONet (black-box)	625k	8.3e−2	9.5e−2	8.6e−2	8.3e−2
DeepSet (black-box)	117k	1.7e−1	1.7e−1	1.7e−1	1.7e−1

Table 3: Interior Laplace on unseen geometries (rel-L_2 vs. exact). The *trained* full NBIO (9k parameters, 3 seeds) and the training-free *structure-only* limit (the 9k correction zeroed) both match the classical solver (4.79e−7 and 4.75e−7 vs. BEM 4.44e−7), confirming that on known-kernel Laplace the accuracy comes from the exact BIE structure, not the learned parameters. The *interior* column is the primary 200-geometry in-distribution test (trained n_b=128); the $n_b \in \{64, 128, 256\}$ columns are the separate resolution-transfer sweep, so the two n_b=128 entries measure different query/refinement protocols and need not coincide exactly. Black-box baselines stay near 10^{-1}, and NBIO's error *decreases with boundary resolution* (trained at n_b=128) like a genuine quadrature while they are flat.

Ablation (where value comes from). Black-box DeepONet 8.3e−2 → learned-kernel BIE (no split) 2.3e−3 (35×) → NBIO split+correction (+known-kernel split) 4.8e−7 (+4800×): the architecture (M2/M3) already beats black-box, and the known-kernel split earns solver grade. Direct and Neumann solves agree.

Stronger geometry-aware baselines on the same data. To address the concern that DeepONet/DeepSet are too weak for irregular geometry, we additionally train geometry-aware neural operators on the identical interior-Laplace boundary-to-query dataset. A faithful GeoFNO implementation is the strongest such baseline, reaching 1.21e−2 ± 3.3e−4 over three seeds (best 1.17e−2), improving substantially over DeepONet and the Boundary Transformer but still remaining $\sim 2.5 \times 10^4$ above NBIO's error. Canonical GNO from `neuraloperator` reaches only 2.62e−1 ± 2.7e−3 under a capped but full-test protocol; canonical GINO instantiated but was not scored because the library path was prohibitively slow for per-sample changing boundary geometries. Thus the conclusion is not an artifact of only comparing to simple black-box baselines: geometry-aware volumetric/graph operators help, but still do not inherit quadrature convergence or reference-solver accuracy.

method	params	interior-test rel-L_2	provenance
NBIO	**9k**	**4.8e−7**	BIE layer, same data
GeoFNO	6.6M	1.21e−2 ± 3.3e−4	faithful pure-torch implementation, 3 seeds
GNO	211k	2.62e−1 ± 2.7e−3	`neuraloperator` v2.0.0, capped training/full test
Boundary Transformer	878k	2.38e−2	same-data boundary attention baseline
DeepONet	625k	8.11e−2	same-data black-box baseline
DeepSet	117k	1.66e−1	same-data black-box baseline

Table 4: Stronger geometry-aware baselines on the same interior-Laplace task. GeoFNO is the strongest non-BIE baseline and materially improves over DeepONet/DeepSet, but remains orders of magnitude above NBIO. The DeepONet/DeepSet entries (8.11e−2, 1.66e−1) are this study's same-data re-evaluation and agree with the interior-accuracy table (8.3e−2, 1.7e−1) to within run-to-run variation.

The conclusion also holds across test instances, not only after averaging or across training seeds (Table 5). On the frozen 200-geometry test set, NBIO's median and p90 track BEM, whereas the no-split and GeoFNO distributions remain separated by orders of magnitude.

method	mean [95% bootstrap CI]	median	p90	p99
classical BEM	4.44e−7 [3.05, 6.27]e−7	1.43e−7	9.61e−7	4.60e−6
NBIO split+corr. (3 seeds)	4.79e−7 [3.35, 6.71]e−7	1.81e−7	8.83e−7	4.87e−6
learned-kernel BIE (3 seeds)	2.35e−3 [2.22, 2.51]e−3	2.09e−3	3.46e−3	5.72e−3
GeoFNO (3 seeds)	1.21e−2 [1.14, 1.29]e−2	1.07e−2	1.86e−2	3.44e−2

Table 5: Test-instance uncertainty on the frozen interior-Laplace test set ($N = 200$). For learned models, predictions are averaged by instance across three seeds before computing distributional statistics and paired-instance bootstrap intervals. Seed dispersion is reported separately in Tables 4 and 9.

4.2 Does the learned part do real work? Identifying a missing screening law

Motivating failure of a generic correction. For $(\Delta - \kappa^2)u = 0$ at fixed $\kappa = 4$, a deliberately wrong Laplace BIE gives rel-$L_2 = 1.13$. A generic Laplace-prior correction reduces this to 8.27e−3±3.7e−4, but a freely learned BIE kernel is better (2.85e−3±3.1e−4). Thus a globally wrong prior can constrain an additive correction. This experiment establishes that learning removes discrepancy, but not that the original correction is the right mechanism.

Proposition 3 (non-identifiability without the operator parameter)**.** *For fixed nonzero Dirichlet data, solutions of $(\Delta - \kappa_i^2)u = 0$ at distinct $\kappa_1 \neq \kappa_2$ differ. Consequently a deterministic model that does not receive κ cannot represent both, and its maximum error is at least half their separation.*

This follows by subtracting the equations; equality would force $u \equiv 0$. We therefore supply κ and test the radial normal form in Eq. (1). This is controlled identification of a screened family, not an "unknown physics" claim.

Parameter-matched ablation. We train three paired seeds with identical 7,500-update budgets and select checkpoints using validation data before held-out evaluation. The test set crosses four geometry families, point-source/Fourier loads, five unseen $\kappa \in [0, 4]$, and $n_b \in \{64, 128, 256\}$. Table 6 reports the primary aggregate at $n_b = 128$; Appendix E records the checkpoint-sealing and test-access protocol.

This ablation isolates two effects. Removing angular pair features is useful, but insufficient: the separate-input radial model remains 28.4× worse than the dimensionless normal form. Capacity does not explain the result because all learned controls differ by only 20 parameters. The independent prior-free control is also worse, so the gain is not an artifact of forcing the generic model to retain a globally wrong Laplace prior. The radial model outperforms the generic controls in every evaluated geometry, load, and parameter cell; the maximum radial/exact condition-number ratio is 1.002 and maximum solve residual is 2.58e−7.

method	parameters	registered holdout	independent replicate
exact modified-Helmholtz BIE	0	6.98e−5	4.30e−5
generic Laplace-prior correction	2,221	1.18e−2	1.08e−2
generic prior-free learned kernel	2,221	—	1.62e−2
radial inputs (κ, r)	2,241	4.28e−3	4.50e−3
radial normal form κr	2,241	**1.51e−4**	**1.33e−4**

Table 6: Mean per-instance rel-L_2 at $n_b = 128$. The prospective registered holdout has 128 geometries, two loads and five unseen κ values. Its normal-form/full-prior ratio is 0.0128 (hierarchical 95% interval $[0.0114, 0.0143]$), and its normal-form/separate-input ratio is 0.0353 ($[0.0325, 0.0382]$). The bootstrap resamples training seed and geometry while retaining each geometry's loads and parameter values. A second independently seeded, checkpoint-sealed holdout replicates the radial result and adds the matched prior-free control; its radial/full-prior and radial/prior-free ratios are 0.0123 and 0.00820. Every paired radial seed beats each available generic control.

Interpolation, covariance, and recovered law. A separate prospective sparse-parameter run trains only at $\kappa \in \{0, 1, 2, 4\}$ and tests seven unseen interior values. Mean error is 2.84e−4 for the normal form, 5.77e−3 for the best generic control, and 1.10e−4 for the exact BIE. The generic prior is slightly better at the smallest value $\kappa = 0.25$ (6.40e−4 versus 7.69e−4). Under $x' = ax, \kappa' = \kappa/a$, normal-form equivariance drift is at most 1.57e−7, whereas generic controls drift by up to order one, consistent with Prop. 1. Post-hoc, the learned multiplier follows the withheld $s\mathrm{K}_1(s)$ law over densely sampled separations but not in the rare tail (Fig. 17); we therefore claim interpolation on the data-supported range, not unrestricted extrapolation.

Contiguous interval holdout and amortized baseline. We next remove the complete interval $(1.5, 2.5)$ from training, rather than isolated parameter values. Training uses equal numbers of continuous samples from $[0, 1.5]$ and $[2.5, 4]$; testing uses five values in the missing interval, four fresh geometry families, two load families, and three resolutions. Table 7A shows that the normal form remains close to the exact-kernel discretization floor and is 17.9× more accurate than the stronger separate-input control. All three paired seeds win, and the geometry-cluster bootstrap interval for the error ratio is $[0.031, 0.102]$.

For a non-BIE comparison, we train a 878k-parameter κ-conditioned Boundary Transformer on the identical training and holdout protocol. Its targets and quadrature weights are normalized using training-only statistics. On its fresh holdout, the Transformer obtains 2.88e−2, versus 1.56e−4 for the radial model (Table 7B). These methods have different contracts: the Transformer is an amortized forward pass, whereas radial NBIO assembles and solves a new system. At batch size one on the same Titan X GPU, their measured medians are 3.30 and 2.35 ms/sample, respectively; this is not a claim about throughput under optimal batching.

method	interval holdout (A)	Transformer holdout (B)
exact modified-Helmholtz BIE	5.95e−5	6.91e−5
generic Laplace-prior correction	6.29e−3	7.31e−3
generic prior-free learned kernel	—	5.74e−3
radial inputs (κ, r)	1.91e−3	—
radial normal form κr	**1.07e−4**	**1.56e−4**
κ-conditioned Boundary Transformer	—	2.88e−2

Table 7: Mean per-instance rel-L_2 at $n_b = 128$. (A) Prospective contiguous parameter-interval holdout; all learned BIE rows use three paired seeds. (B) Fresh same-protocol amortized baseline; all learned rows use three seeds. The two panels use different fresh holdouts and are not compared across columns.

Solution-observation efficiency. With optimizer updates fixed at 7,500, we retrain the radial and matched generic models on deterministic subsets of the original training distribution. Table 8 shows that the radial law is already identified from 100 unique solution instances. A separate 300-instance pilot uses 32 query observations per instance and fixed iid noise scaled to each solution's RMS. At 5% noise the radial model degrades to 1.20e−3 but remains 13.8× more accurate than the generic correction. The 100-instance ra-

dial/generic ratio has clustered 95% interval $[0.0109, 0.0134]$; the 5%-noise ratio has interval $[0.0667, 0.0782]$. This establishes efficiency in unique supervised observations at fixed training compute, not reduced optimizer cost or robustness to arbitrary sensor noise.

unique instances	radial	generic	32-query noise	radial	generic
100	1.23e−4	1.02e−2	0%	1.48e−4	1.12e−2
300	1.39e−4	9.79e−3	1%	2.91e−4	1.25e−2
800	1.32e−4	9.42e−3	5%	1.20e−3	1.65e−2
2400	1.34e−4	9.76e−3	—	—	—

Table 8: Clean held-out mean rel-L_2 for the pre-specified data-efficiency grid evaluated on an existing held-out set. Left: 256 query observations per training instance. Right: 300 unique instances and 32 observations each with synthetic relative label noise. All rows use three paired seeds and the same 240,000 examples seen through recycling; checkpoint selection uses a clean validation set.

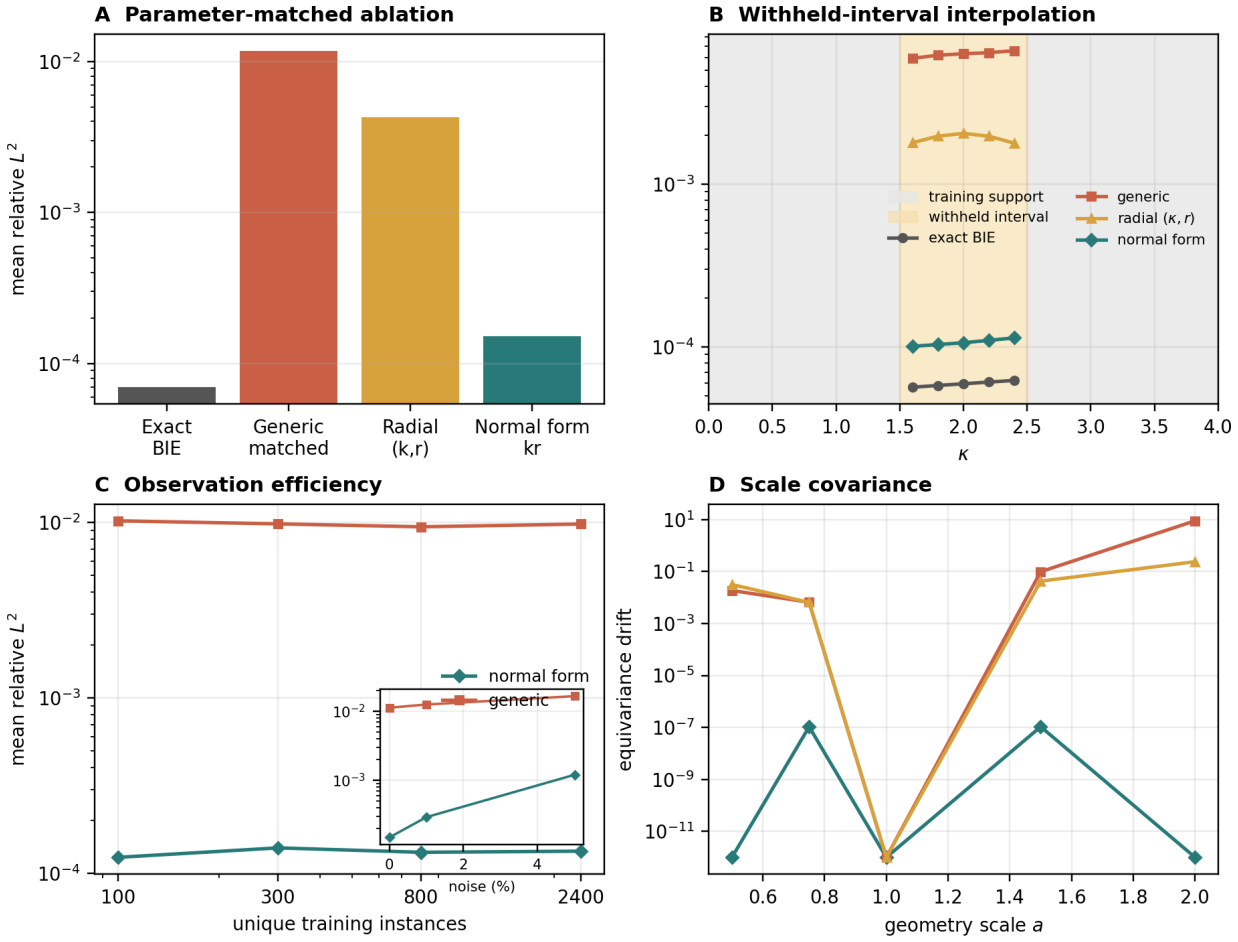

Figure 1: Learned-discrepancy evidence. (A) Parameter-matched ablation. (B) Interpolation across a fully withheld parameter interval; gray denotes training support and yellow the excluded interval. (C) Unique-observation efficiency at fixed training compute, with the 32-query synthetic noise study inset. (D) Joint geometry/parameter scale covariance.

Near-boundary evaluation. For queries at distances 10^{-2}–5×10^{-4}, direct trapezoidal evaluation fails for both the exact and learned kernels (pooled error ≈ 7.07). We split $K_0 m_\theta = K_0 + K_0(m_\theta - 1)$, apply differentiable QBX only to the retained singular K_0 term, and evaluate the learned remainder directly. On

fresh in-distribution and rough geometries this gives 7.20e−5 for exact-kernel split-QBX and 8.32e−4 for radial split-QBX; query and parameter gradients are finite and nonzero. The repair removes the direct blow-up, but the radial model is 11.6× above the exact-kernel discretization floor, so we do not call this solver-grade learned close evaluation.

Differentiable and computational utility. Across four unseen Fourier-shape targets and four fixed starts each, the radial model succeeds in all 16 runs with mean coefficient RMSE 2.48e−5, versus 3.51e−4 for the better generic control and 1.84e−5 for the exact kernel. Its minimum initial-gradient cosine to the exact model is 0.99999988. Figure 2 shows the coefficient-recovery comparison. Runtime against the amortized Transformer is reported above under one unified batch-size-one harness; the distinct component-timing audit is deferred to Appendix A.

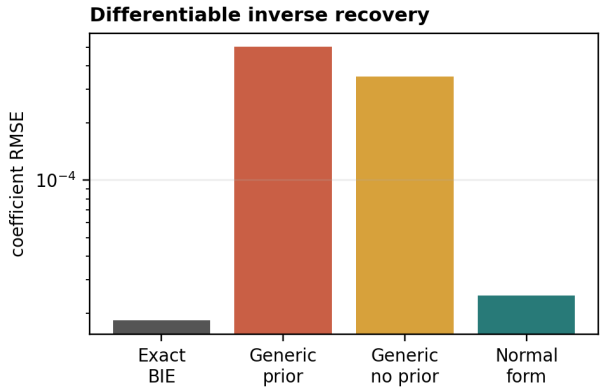

Figure 2: Four-coefficient inverse recovery through the differentiable solve. Bars show mean coefficient RMSE over the fixed-start runs.

The earlier equal-capacity generic κ-conditioned experiment remains in Appendix B as a secondary control. Its mixed results motivated the radial factorization; it is no longer the paper's primary learning claim. The practical rule is to retain exact singular structure, parameterize missing physics in the smallest symmetry- and dimension-consistent family available, and abandon a prior when a frozen no-prior control wins.

Ablation (M1/M2/M3, interior Laplace, 3 seeds). Each ingredient is isolated below (Table 9): dropping the known-kernel split (M1) costs ∼4 decades; the iterative second-kind solve (M2) matches the direct solve; and at zero learned correction NBIO is the classical solver (training-free 4.75e−7), consistent with Prop. 2.

variant (interior Laplace)	params	test rel-L_2 (mean±std)
NBIO split+corr. (M1+M2+M3)	9,026	**4.79e−7** ± 1.5e−10
−M1: no known-kernel split (learned kernel)	9,025	2.35e−3 ± 1.2e−4
M2 iterative (damped Neumann vs direct)	9,026	5.03e−7 ± 1.1e−8
no learned correction (classical, training-free)	0 active	4.75e−7
classical Nyström BEM (reference)	—	4.44e−7

Table 9: Ablation of the three ingredients. M1 (known-kernel split) is worth ∼4 decades; M2 direct≈iterative; the learned correction is unnecessary on known-kernel Laplace (NBIO reduces to the classical solver). Table 6 separately isolates learning when the screened constitutive law is withheld.

Related volumetric control: variable principal coefficient. For $\nabla\cdot(a(x)\nabla u) = f$ there is no closed-form Green's function. This experiment is not the boundary architecture above: it supplies a constant-coefficient FFT solution as an extra channel and adds a four-layer FNO residual on a periodic grid,

$$u_\theta \;=\; u_{\text{const}} \;+\; \mathcal{F}_\theta(a,\, f,\, g,\, u_{\text{const}})\,, \tag{2}$$

where u_{const} is the constant-coefficient FFT prior and $\mathcal{F}_\theta$ is the four-layer residual FNO on the periodic grid. This is a volumetric structured-prior control, not an instance of the boundary-integral NBIO architecture

defined in Section 3. We include it to test the broader "known prior plus learned discrepancy" principle. The fixed prior gives 9.0e−2; the prior-plus-residual model reaches 3.4e−3±1.5e−4 over 5 seeds vs. a black-box FNO 4.3e−3±1.0e−4. *Scope:* the learned correction is *necessary* (∼27× over the fixed prior), but on this volume problem the structural edge over a strong FNO is modest (∼1.3×) and contrast extrapolation is essentially tied – the boundary/Green's advantage that dominates BIE-native problems largely evaporates here. *On the recognized released benchmark* (FNO Darcy tensors, identical data to both models, 3 seeds) the structure-informed prior in fact does not beat the black-box in-distribution (7.6e−2 vs 6.6e−2 at res 32; a ∼9% edge only on zero-shot super-resolution), reinforcing (on external footing) that this volumetric high-contrast regime is outside NBIO's structural niche.

4.3 Out-of-distribution generalization

A learned solution map is only as useful as its behaviour off the training distribution. In this subsection "NBIO" denotes the trained full model, but its known-kernel accuracy is structure-dominated, as established by Table 9. Training NBIO on the 5-mode smooth-star family and evaluating zero-shot (same n_b, so this isolates geometry shift, not resolution) on harder families, NBIO tracks the classical BEM reference to within ∼3% on every shift, while a black-box DeepONet stays collapsed at $\sim 10^{-1}$ (Table 10, Fig. 3). NBIO's behaviour is consistent with a *discretized operator* rather than a fitted map: its OOD error is the under-resolved-quadrature error the classical solver also incurs on rougher boundaries, and it is removed zero-shot by raising n_b – consistent with Prop. 2 applied separately on each fixed geometry, not with a family-uniform theorem.

rel-L_2	in-dist	high-mode	rough	deformed	near-bdry
classical BEM (ref)	4.4e−7	1.1e−5	8.3e−5	7.7e−5	1.8e−4
NBIO split+corr. (9k; structure-dominated)	**4.8e−7**	1.2e−5	8.2e−5	8.0e−5	1.8e−4
DeepONet (625k)	8.3e−2	8.8e−2	1.1e−1	9.7e−2	8.7e−2

Table 10: Zero-shot out-of-distribution generalization (interior Laplace, trained only on the 5-mode smooth-star family). The trained full NBIO tracks the classical BEM reference within ∼3% on every shifted family because its exact-kernel result is structure-dominated; this table is structural transfer evidence, whereas Table 6 isolates transfer of a load-bearing learned law.

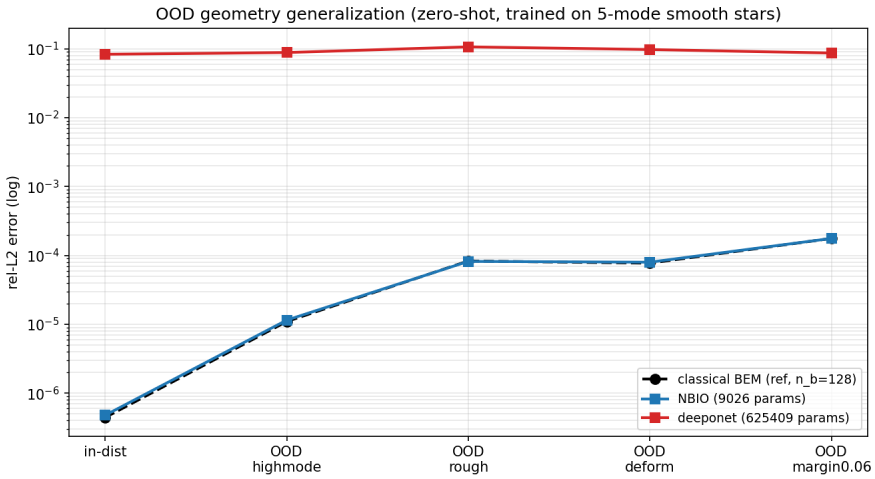

Figure 3: Out-of-distribution generalization: NBIO (trained in-distribution) tracks the classical BEM reference across geometry-distribution shifts, 3–4 orders below the black-box.

4.4 Exterior, transmission, nonsmooth geometry

Exterior (unbounded). NBIO 6.2e−7 ≈ deflated exterior BEM 5.6e−7; black-box 0.12–0.46 – a capability volumetric methods do not natively represent without truncation, mapping, or specialized treatment. **Transmission** (piecewise-constant a, inclusions): the NBIO-transmission *structure* (zero trainable parameters = the classical multi-region solver) holds ∼1.3e−6 across unseen geometries, inclusion counts (1/3/4, trained-equivalent at N=2), held-out high contrast ($\sigma \in [6, 12]$), and all resolutions; DeepSet 0.14–0.58. **Nonsmooth families**: NBIO tracks the BEM reference on high-curvature (1.6e−5), rounded-triangle, peanut (1.0e−7), and rough boundaries; the stress case is true near-corners (reference itself only 2.7e−4).

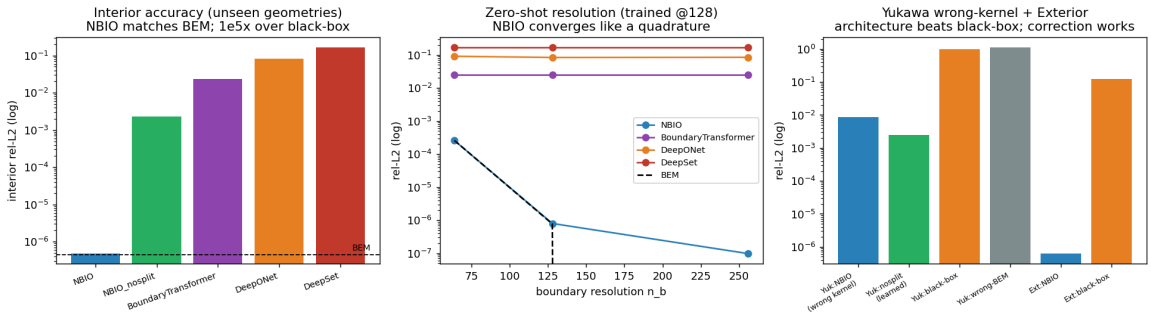

Figure 4: Interior accuracy, zero-shot resolution, and wrong-kernel/exterior results. The known-kernel interior, resolution, and exterior gains are structure-dominated; the modified-Helmholtz row is the motivating fixed-parameter generic-mismatch baseline. The primary learning result is the radial normal-form experiment in Fig. 1 and Table 6.

5 Extensions and numerical integration studies

The preceding experiments separate three findings. Exact BIE structure supplies solver consistency in the known-kernel regime. A generic pairwise correction can repair fixed kernel mismatch but may remain biased by a globally inaccurate prior. For the screened family, imposing the radial normal form and learning only a dimensionless law of κr gives the primary learning result, with measured transfer across parameters, geometry, loads, discretizations, and scale. The remaining studies test how the resulting layer behaves on real meshes, close evaluation, preliminary 3D, interfaces and sources, structured acceleration, and embedding in nonlinear or transient algorithms. They are extensions of the core claim rather than separate claims of universal neural-operator dominance.

5.1 Real CAD geometry: exterior electrostatics on ABC and Fusion 360

The results above use analytic or manufactured geometry families. To test the representation on real, uncurated engineering geometry, we evaluate exterior Laplace electrostatics (potential from prescribed boundary voltages) on watertight meshes from the ABC and Fusion 360 CAD corpora. Here the NBIO/BIE object is the *zero-parameter analytic structure* – the exact single-layer exterior solve, no trainable parameters – exactly the physics fallback used in the transmission and interface tables. On a frozen protocol (fixed meshes, voltage families, train/test split, and evaluation subsampling), the structure reaches median rel-$L_2 = 2.08 \times 10^{-3}$ on ABC (n=145) and 5.56×10^{-3} on Fusion (n=95). As strong learned points of comparison on the *identical* frozen protocol we train geometry-aware neural operators (spatial-bias attention on ABC, a neural Green's-function operator NGF on Fusion) and re-run them over three model-initialization seeds. Mesh-disjoint validation selects each checkpoint; the held-out test meshes are not evaluated until selection is complete. Their seed-averaged median rel-L_2 is 0.522 ± 0.020 (ABC) and 0.528 ± 0.042 (Fusion), i.e. 251× and 95× above the structure (Table 11, Fig. 5). The gap is stable to model initialization. These leakage-free reruns supersede an earlier internal aggregate whose trainer selected checkpoints on a test subset; no number from that aggregate is used here.

Does learning help on known-kernel CAD? No. Because the exterior Laplace kernel is known, this is the CAD analogue of the interior-Laplace ablation: the split already supplies the accurate structural solve, so a learned correction has nothing to add. We test this directly with a paired, identical-case residual: on Fusion a learned BIE-residual correction worsens the coarse-BIE prior (median 5.03e−3 → 6.20e−3, p90 1.23e−2 → 1.32e−2), and on ABC it improves the median by only 4.96% (3.91e−3 → 3.71e−3) while worsening the mean, p90, and maximum. We therefore *do not* claim that learning improves the known-kernel CAD solver; the real-CAD result is evidence for the boundary-integral *representation*, not for a learned-model CAD state of the art. This mirrors the interior-Laplace finding (Table 9): the learned correction is necessary only where the kernel is wrong or unknown.

corpus	method	median rel-L_2	vs. struct.	provenance
ABC	**NBIO/BIE structure**	**2.08e−3**	—	0-param exterior solve
	spatial-bias attention	5.22e−1 ± 2.0e−2	251×	learned, 3 seeds
Fusion 360	**NBIO/BIE structure**	**5.56e−3**	—	0-param exterior solve
	neural Green's operator (NGF)	5.28e−1 ± 4.2e−2	95×	learned, 3 seeds

Table 11: Real CAD exterior electrostatics (frozen, leakage-free protocol). Values for learned models are the mean and sample standard deviation across three seed medians. The zero-parameter BIE structure is 95–251× below the geometry-aware learned operators. This is real-geometry evidence for the *representation*; the learned CAD residual does not improve the known-kernel solver (paired test, text) and no learned-CAD SOTA is claimed.

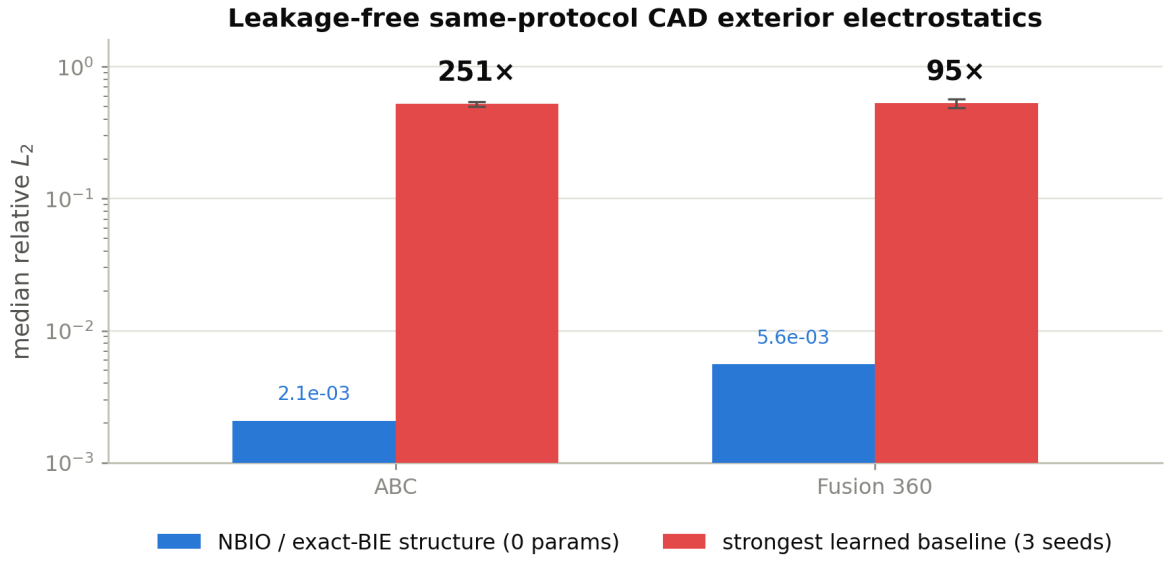

Figure 5: Real CAD geometry (ABC and Fusion 360 exterior electrostatics). The zero-parameter NBIO/BIE structure is compared with the strongest tested learned geometry operator on each identical mesh-disjoint protocol. Error bars show sample standard deviation across three seed medians; the learned models remain 95–251× above the structure.

5.2 Close evaluation at near-touching interfaces (differentiable QBX)

Near the boundary, and especially in the thin gap between near-touching inclusions, layer potentials are near-singular and naive quadrature fails (both NBIO and black-box reach rel-$L_2 \in [10^{-1}, 10^{0}]$). This is a *close-evaluation*, not a representation, failure. We add a differentiable QBX (Klöckner et al., 2013) that evaluates the known (gated) kernel by a local harmonic expansion off the surface; the nonsingular learned correction stays a direct evaluation. *In-operator*, on the trained but known-kernel/structure-dominated

model: away-from-boundary accuracy is retained (3.2e−7), zero-shot resolution is retained, and near the boundary the error drops from 13.5 (direct) to 3.9e−6 (QBX) at distance 10^{-3} (a $\sim 10^6\times$ improvement), while the operator remains differentiable end-to-end. For near-touching inclusions, QBX evaluation plus oversampled/adaptive cross-block assembly recovers the gap-region error in the transmission operator (e.g. gap 0.005: 0.65 → 1.5e−3, ∼430×; Fig. 10 left).

We validate this implementation claim numerically in float64 using central finite differences at $\epsilon = 10^{-5}$ (also swept from 10^{-3} to 10^{-5}). Directional derivatives through the boundary datum, a learned correction parameter, a Fourier geometry coefficient, and a near-query QBX coordinate agree with automatic differentiation to relative discrepancies 4.2e−15, 9.5e−12, 7.8e−9, and 1.5e−9, respectively. The QBX nearest-node choice is unchanged under the perturbation; as with other nearest-neighbour branches, differentiability is piecewise at selection boundaries.

derivative path	relative AD–FD discrepancy	modules traversed
boundary datum g	4.2e−15	solve → potential
learned correction parameter	9.5e−12	assembly → solve → potential
geometry Fourier coefficient	7.8e−9	geometry/weights → assembly → solve → potential
near-query coordinate	1.5e−9	QBX potential evaluation

Table 12: End-to-end directional gradient checks. All four discrepancies are below the pre-registered 10^{-3} threshold.

Gradient-based use. To test more than local derivative agreement, we recover four Fourier geometry coefficients from 64 interior observations under four boundary loads at $\kappa = 2.2$. Targets come from a refined 512-node correct-Yukawa BIE; optimization differentiates through a 128-node solve. Across eight fixed starts, the smooth full-prior model reaches coefficient RMSE 3.34e−4 and observation error 2.96e−3; the smooth no-prior model reaches 8.97e−5 and 1.93e−3. Both pass the predeclared thresholds for all starts. The no-prior advantage is retained: this demonstrates usable end-to-end gradients, not a prior-specific inverse-problem win. The stronger radial-normal-form inverse study in Fig. 2 uses four unseen targets and reports 2.48e−5 mean coefficient RMSE over 16 fixed-start runs.

5.3 Comparison to prior boundary/interface operators (same protocol where possible)

External operators differ in PDE, supervision, and reference solver, so a single shared leaderboard is not meaningful. We therefore separate three cases: a matched problem-specification comparison against a published comparator, where we reproduce the geometry and manufactured solution but quote the published error rather than rerun its code (LOB); controlled same-protocol interface reproductions where we match the relevant input class (contrast, jumps, and volume source), and broader related-work positioning where the released protocols target different domains or outputs.

Learning-Only-on-Boundaries (2D parametric Laplace). We reproduce the LOB geometry family and its harmonic manufactured solution $u = e^x \sin y$, and quote LOB's reported error (Fang et al., 2024). On this matched geometry and manufactured solution (Table 13), direct Nyström/NBIO is *under-resolved* on the high-curvature boundary at n_b=128, but with the differentiable QBX close-evaluation path NBIO reaches solver grade by n_b=256 – about five orders of magnitude below LOB's reported $\approx 3\times 10^{-2}$ – without retraining on LOB data.

Caveat on exact public protocols. On XI-DeepONet's circular high-contrast *geometry and contrast*, the original NBIO-tm structure reaches $\sim 2\times 10^{-6}$, but that row is only a *homogeneous-transmission subset* and should not be reported as a full XI/IONet benchmark win. Below, we close the input-class gap with interface jumps and volume sources. The black-box baselines are trained on our faithful reproduction of that problem class, while IONet/XI/BENO numbers are published reference points; we do not claim to have rerun their exact released datasets.

Interface problems with jumps (controlled same-protocol comparison). We extend NBIO to ingest prescribed interface jumps $[u] = \alpha$, $[a\,\partial_n u] = \beta$ through a direct (Calderón) formulation, validated to machine precision ($\sim 10^{-15}$) against manufactured piecewise-harmonic solutions on circular and irregular

method (geometry t=1.15)	n_b=128	n_b=256	n_b=1024
Published LOB (PINO-BIE) (Fang et al., 2024)	2.85e−2 (single published value)		
NBIO (direct evaluation)	2.22e−1	7.70e−2	4.40e−4
NBIO + QBX	1.31e−2	**6.9**e−**7**	**5.5**e−**7**
DeepONet (black-box)	4.37e−1	4.64e−1	4.55e−1

Table 13: Matched problem-specification comparison on the LOB 2D parametric Laplace problem (rel-L_2; results for t=1.35, 1.45 are similar). NBIO with QBX close evaluation reaches solver grade by n_b=256, vs the published $\approx 3\%$. The "Published" row quotes (Fang et al., 2024); NBIO rows are our evaluation on a reproduction of their geometry and manufactured solution.

interfaces across contrasts. On a parametric family of *source-free* elliptic interface problems (varying interface geometry, contrast 2–8, jump and boundary data), NBIO-interface reaches solver grade zero-shot (median rel-$L_2 = 1.5 \times 10^{-15}$, mean 3.3×10^{-12}, 0 trainable parameters); strong black-box operators trained on the same data reach 6.8×10^{-2} (cross-attention Transformer) and 2.2×10^{-1} (DeepSet), against which NBIO is many orders of magnitude more accurate where the interface reduces to known-kernel structure. The purpose-built interface operators IONet (Wu et al., 2024) and XI-DeepONet (Bi et al., 2025) report $\sim$3–5×10^{-3} on related interface problem classes, but on different datasets and protocols, so we treat them only as contextual reference points (Table 14). *Scope:* this first interface table is source-free; the next table adds the full inhomogeneous source term. Black-box rows use a faithful reproduction of the problem class (not the exact released datasets); IONet/XI figures are quoted for context.

parametric elliptic interface (jumps, source-free)	params	rel-L_2
NBIO-interface (structure, zero-shot)	**0**	**3.3**e−**12** (med. 1.5e−15)
same-protocol (our reproduction, identical data/splits):		
Interface Transformer	0.30M	6.8e−2
Interface DeepSet	0.40M	2.2e−1
published context (different datasets/protocols; not head-to-head):		
IONet (Wu et al., 2024)	—	$\sim$3.5e−3
XI-DeepONet (Bi et al., 2025)	—	$\sim$3–5e−3

Table 14: Controlled same-protocol comparison on the source-free elliptic interface problem with jumps. NBIO-interface (direct Calderón solve with $[u], [a\partial_n u]$) matches the reference solver zero-shot; trained black-box operators and published interface-specialized operators remain at 10^{-3}–10^{-1}.

Full inhomogeneous problem (volume source + jumps). This is a hybrid volume–boundary method, not a boundary-only computation. The particular-solution split reduces the inhomogeneous case to the jump solver; with bicubic FFT-Poisson interpolation on a 1024^2 grid, the manufactured-solution validation reaches $\sim 1.4 \times 10^{-9}$. On the full family (contrast, prescribed jumps, and a volume source), NBIO-interface-source reaches mean rel-$L_2 = 3.1 \times 10^{-8}$ (median 1.9×10^{-9}) zero-shot with 0 trainable parameters (Table 15, Fig. 6). Against the *same-protocol* source-aware black-box operators trained on the identical data (4.2×10^{-1} and 5.2×10^{-1}), NBIO is about seven orders of magnitude more accurate; the published IONet ($\sim 1.7 \times 10^{-2}$) and XI-DeepONet ($\sim$ 3–5×10^{-3}) are cited only as context (different datasets and protocols, so we do not compute a cross-dataset ratio). This accuracy is achieved by upgrading the volume-potential interpolation from bilinear to bicubic. The FFT grid occupies approximately 8 MiB per real float64 field before work arrays (complex FFT arrays and coordinates increase peak memory), uses a periodic box $[-2.8, 2.8]^2$, subtracts the box mean of f, and relies on compact sources plus padding so periodic-image effects are negligible. Bicubic interpolation is the measured error floor. It is independent of the boundary mesh but is neither mesh-free nor suitable for arbitrary non-periodic sources without a different volume quadrature.

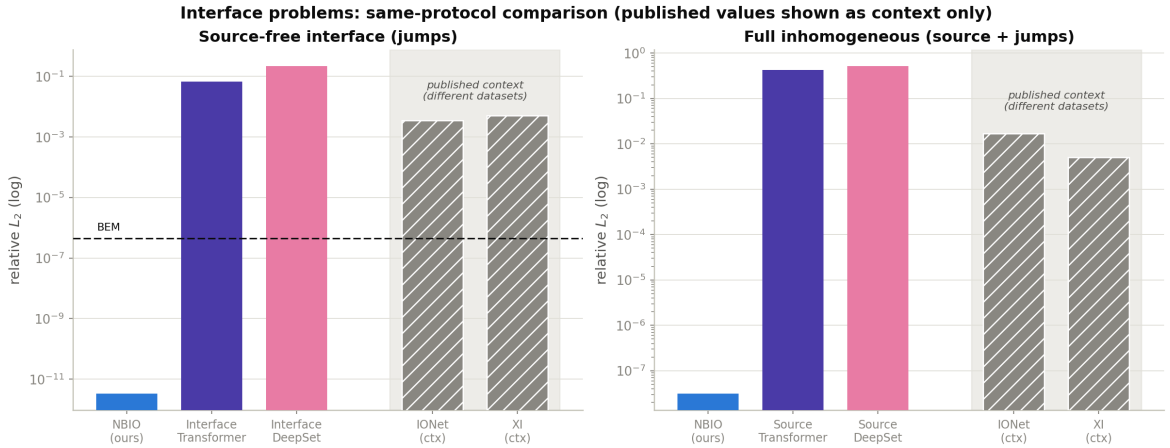

Figure 6: Same-protocol comparison on the elliptic interface problem family. Left: source-free (jumps only). Right: full inhomogeneous (contrast, jumps, volume source). NBIO is zero-shot (0 params) and black-box baselines are trained on the same generated data; the published IONet/XI values sit in the *shaded context region* (different datasets/protocols) and are not directly comparable.

inhomogeneous interface (source + jumps)	params	rel-L_2
NBIO-interface-source (structure, zero-shot)	**0**	**3.1**e−**8** (med. 1.9e−9)
same-protocol (our reproduction, identical data/splits):		
Source Transformer	0.30M	4.2e−1
Source DeepSet	0.40M	5.2e−1
published context (different datasets/protocols; not head-to-head):		
IONet (Wu et al., 2024) (2D source+boundary)	—	∼1.7e−2
XI-DeepONet (Bi et al., 2025) (interface class)	—	∼3–5e−3

Table 15: Controlled same-protocol comparison on the inhomogeneous elliptic interface problem with contrast, prescribed jumps, and a volume source. NBIO closes the source/jump input-class gap; the error floor is currently set by the FFT/interpolation particular solution rather than by the boundary solve. Published rows are quoted reference points, not reruns of released datasets.

5.4 Scalability: are the $O(n^2)$ assembly and $O(n^3)$ solve removable?

A standing concern for any dense BIE method is cost: assembling the kernel matrix is $O(n_b^2)$ and the direct solve is $O(n_b^3)$. We give direct evidence that both costs are removable *without loss of accuracy* on the interior-Laplace problem (Fig. 8); the two mechanisms that trade accuracy for speed are reported as limits. For reference we label four scalability techniques **(3)**–**(6)**, distinct from the core modeling components M1–M3 of §3: **(3)** an approximately constant-iteration Krylov solve, **(4)** a separable linear-cost correction, **(5)** adaptive-panel refinement, and **(6)** distillation; (5)–(6) are the accuracy-for-speed limits.

The solve needs an approximately constant iteration count (Method 3). Because the assembled system is a well-conditioned second-kind Fredholm equation, GMRES converges to 10^{-10} in a *constant* ∼8 iterations as n_b grows from 128 to 4096 ($\mathrm{cond}(A) \approx 2.5$ throughout), already beating the dense direct solve by 27× at $n_b = 4096$. The off-diagonal blocks of A moreover have *constant* numerical rank (∼8 at $\epsilon = 10^{-8}$; compression 7× → 128× as the block grows) – the $\mathcal{H}$-matrix/fast-multipole compressibility that yields an $O(n_b \log n_b)$ matvec. Together these are the two prerequisites for a near-linear fast solver.

A separable linear-cost correction in the known-kernel regime (Method 4). Replacing the pairwise nonsingular correction kernel $K_{\mathrm{smooth}}(x_i, y_j)$ (the K_θ^{corr} term) with a separable $\phi(x_i) \cdot \psi(y_j)$ (per-node MLPs mapping the 5 local geometric features to $\mathbb{R}^r$) makes the correction matvec $O(n_b r)$. At *rank* 4 the

factored model retains full solution accuracy on the *known-kernel* interior-Laplace experiment (4.81e−7 vs 4.79e−7) with unchanged conditioning; the factored matvec is essentially flat in n_b (∼0.06 ms) while the dense correction grows quadratically (394 ms at $n_b = 2048$, a ∼7500× gap), showing the correction path can be made linear-cost in this regime. Because the exact core already reaches solver grade here, this does not establish that a substantial wrong-kernel correction is rank-4; whether similarly low ranks suffice when the correction is load-bearing (§5.4 Helmholtz, wrong-kernel) remains open.

End-to-end near-linear solve (demonstrated). We realize both ingredients in a single solver (Fig. 7): a recursive $\mathcal{H}$-matrix that compresses every admissible (well-separated) block of A directly from the kernel by adaptive cross approximation – never forming the dense matrix – with near-junction blocks kept dense at the leaves (the correct structure for a closed curve), solved by GMRES. Validated against the dense direct solve, the $\mathcal{H}$-matrix density matches it to $\sim 10^{-11}$ for $n_b \leq 4096$; GMRES stays at 7–9 iterations, the storage fraction shrinks like $\log(n_b)/n_b$ ($0.52 \to 0.007$), and setup+matvec+solve scale near-linearly out to $n_b = 65{,}536$, where the dense $O(n_b^3)$ solve is infeasible. This turns the "in-principle" scalability into a validated near-linear NBIO solver.

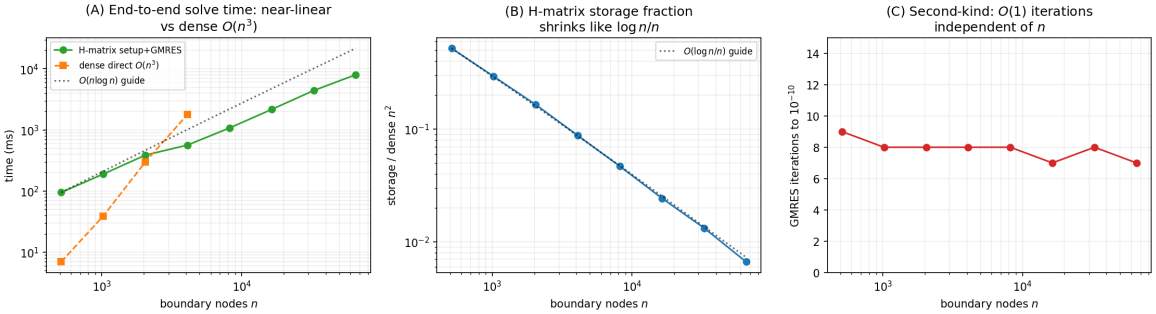

Figure 7: End-to-end near-linear *structure-only* Laplace NBIO/BIE solve (recursive $\mathcal{H}$-matrix + GMRES, interior Laplace). (A) total solve time tracks $O(n \log n)$ and crosses below the dense $O(n^3)$ solve; (B) $\mathcal{H}$-matrix storage shrinks like $\log n/n$; (C) GMRES stays at $O(1)$ (7–9) iterations. Density matches the dense solve to $\sim 10^{-11}$ where the latter is feasible.

Limits (Methods 5–6). Adaptive panel refinement (curvature-graded Gauss–Legendre) helps only in the under-resolved regime on a sharp-lobe star; once spectral convergence engages, uniform global Nyström wins, so adaptive panels matter only for non-smooth geometry. Distilling the inverse operator into a feed-forward $g \to \mu$ network gives 19–98× faster inference but costs all the accuracy (6.8e−2 vs 4.8e−7, with the error *entirely* in the density and the known-kernel evaluation exact), quantifying precisely what the assembled solve buys.

Frequency limit (Helmholtz). The constant-rank / approximately-constant-iteration scalability is a *low-frequency* property; on oscillatory Helmholtz the off-diagonal rank and GMRES count grow ∼linearly with wavenumber, and the classical directional/butterfly-FMM remedy (splitting the far field into $O(k)$ directional cones) restores near-linear scaling. Details and Figs. 15 and 16 are in Appendix A.

5.5 Nonlinear problems: NBIO as a zero-shot linear solve inside an outer iteration

One setting where the structured operator is clearly more accurate than a black box is a problem where *no Green's function exists for the full operator*. We demonstrate this on a semilinear elliptic BVP, $-\Delta u + \kappa u^3 = f$ in Ω with Dirichlet data, solved by Picard iterative linearization: u_{n+1} solves the *linear* inhomogeneous problem $-\Delta u_{n+1} = f - \kappa u_n^3$, each step an inhomogeneous Laplace Dirichlet solve handled by the particular-solution split plus a boundary solve. We drop three solvers into the loop – classical Nyström BEM (reference), structure-dominated NBIO (used zero-shot: trained only on linear Laplace Dirichlet data, never on sources or nonlinear problems), and a black-box DeepONet – against a manufactured solution u^* (Table 16, Fig. 9).

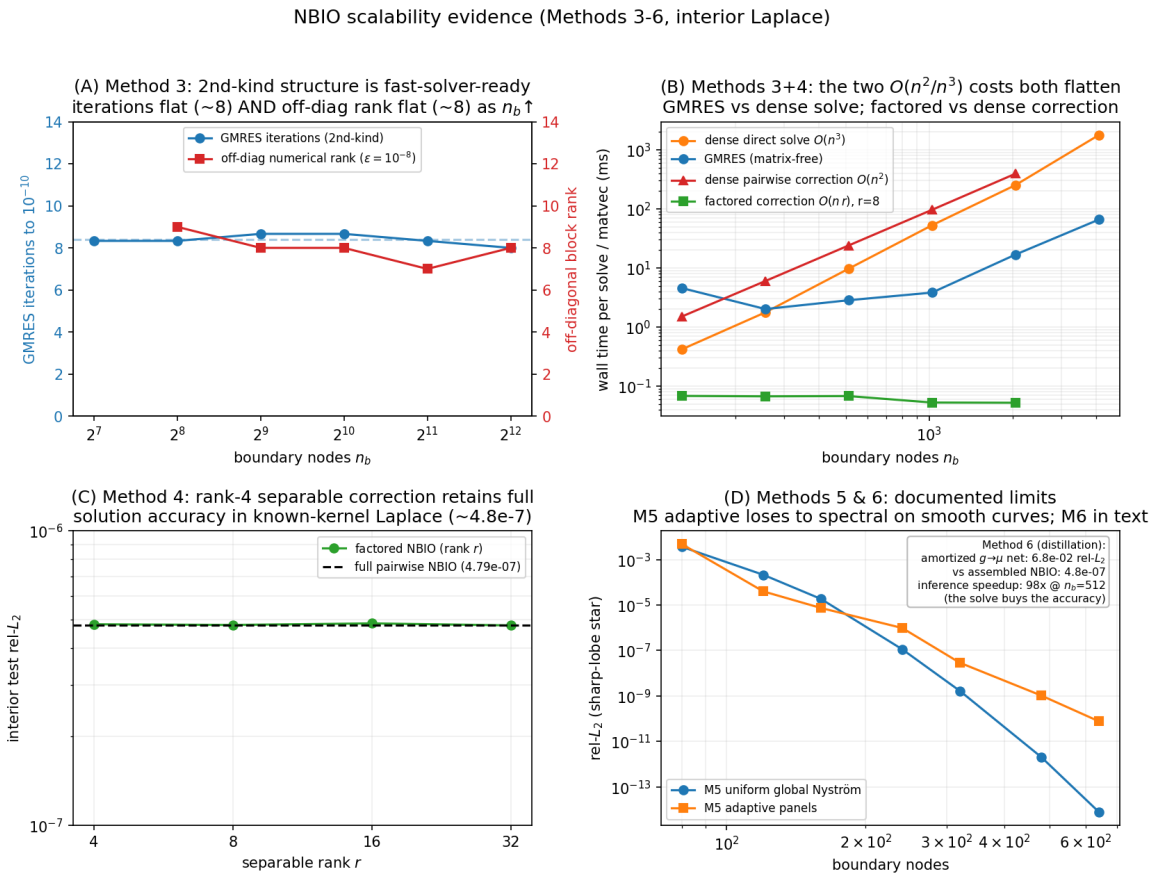

Figure 8: Scalability evidence (Methods 3–6, interior Laplace). (A) Second-kind structure: GMRES iterations and off-diagonal numerical rank are both flat (∼8) as n_b grows. (B) Both $O(n^2/n^3)$ costs flatten: matrix-free GMRES vs the dense solve, and the factored vs dense correction matvec. (C) A rank-4 separable correction retains full solution accuracy in the known-kernel Laplace regime (linear-cost correction path; not a claim about substantial wrong-kernel corrections). (D) Limits: adaptive panels lose to spectral convergence on smooth curves; distillation trades all accuracy for speed (Method 6 annotated).

final rel-L_2 vs u^*	BEM (ref)	**NBIO (structure-dom.)**	DeepONet (black-box)
$\kappa = 0.1$	3.2e−5	**2.9e−5**	1.2e−1
$\kappa = 0.3$	9.9e−5	**8.7e−5**	1.2e−1
$\kappa = 0.6$	2.0e−4	**1.8e−4**	1.2e−1
$\kappa = 1.0$	3.5e−4	**3.0e−4**	1.2e−1

Table 16: Semilinear BVP $-\Delta u + \kappa u^3 = f$ by Picard iteration with each linear step solved by BEM, structure-dominated NBIO (zero-shot), or a black-box DeepONet. NBIO matches the classical reference across the nonlinearity sweep; this is evidence for reusing the exact linear-solver layer, not for a learned nonlinear operator.

NBIO matches the classical-BEM-in-the-loop reference across the entire κ sweep, zero-shot. The black-box DeepONet is pinned at $\sim 1.2 \times 10^{-1}$ (∼400–4000× worse) and converges to the wrong fixed point: its ∼5% per-step error does not shrink as the iterate approaches u^*, so it caps the nonlinear solve. Its reference-solver per-step accuracy allows NBIO to serve as a drop-in linear solve inside the outer nonlinear iteration. Two limitations remain: the shared BEM/NBIO floor ($\sim 10^{-4}$) is set by the FFT particular-solution split, not

the boundary solve (so NBIO is not the bottleneck – it inherits BEM's floor), and this κ range stays in the Picard contraction regime, so the limit here is the linear-solve floor rather than a nonlinearity breakdown.

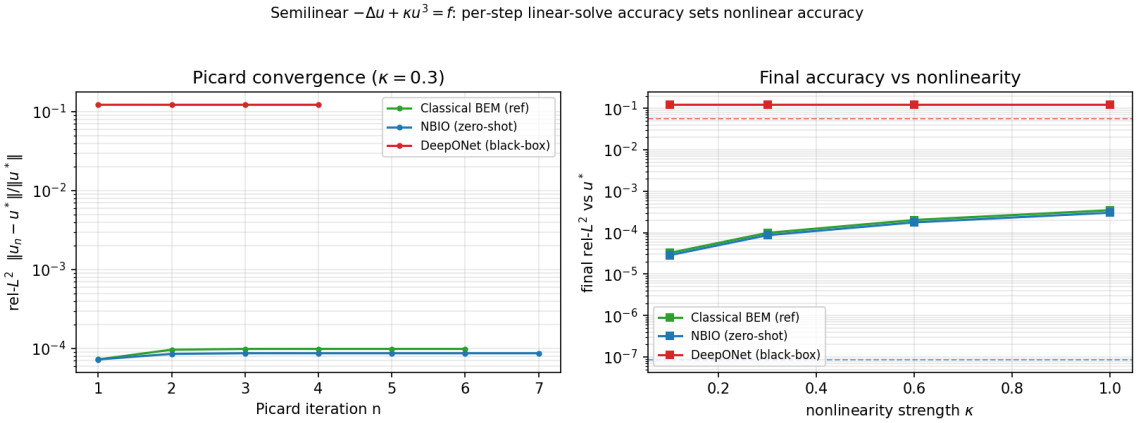

Figure 9: Semilinear BVP by Picard iteration. Left: convergence (rel-L_2 vs iteration) at $\kappa = 0.3$ – zero-shot NBIO (blue) overlaps classical BEM (green) while the black-box DeepONet (red) is pinned at its per-step floor. Right: final accuracy vs nonlinearity strength κ; NBIO tracks BEM, the black box does not.

5.6 Further extensions (deferred to appendix)

Beyond the results above, NBIO supports several further extensions, reported in full in Appendix A: a jointly-trained 3D deformed-surface operator; transient heat solves (implicit-Euler and a causal continuous-time double-layer with arbitrary-time queries); a Newton inner-solve for the semilinear BVP; the variable-coefficient constant-prior Born/GMRES preconditioner; and the high-wavenumber Helmholtz frequency limit with its directional-FMM remedy. Each is structure-consistent and reinforces, rather than extends, the core conclusions.

6 Discussion: scope and limitations

NBIO is strongest where BIE structure matches the operator: interior/exterior/transmission Laplace and interface problems (source-free or inhomogeneous) match or approach reference-solver accuracy without retraining. These 10^2–10^5 same-protocol error ratios are obtained by the *structure-only* BIE limit; they compare a per-instance numerical solve with amortized learned surrogates and should not be read as an orders-of-magnitude gain supplied by neural parameters. Published interface-specialized errors sit on different datasets and are context only. The primary learning evidence is the screened-family experiment in Sec. 4.2: the modified-Helmholtz multiplier is withheld while the analytic singular normal form is retained, and a scalar law of κr is learned from solution supervision. Its transfer is established within the trained parameter range, including a fully withheld interior interval. Although the multiplier is geometry-independent by construction, the interior-trained approximation is not uniformly accurate when inserted into unseen exterior or multi-component integral equations, where changed conditioning amplifies kernel error. The learned-transfer claim is therefore restricted to the simply-connected interior family. The older fixed-κ experiment is a fixed-mismatch baseline: generic correction removes substantial discrepancy, but a globally wrong prior can constrain it. Separately – and *not* using that correction – the structure-dominated exact-Laplace layer is reused without retraining as the per-step linear solve in a Picard iteration, letting NBIO solve a *nonlinear* semilinear BVP to classical-solver accuracy where a black-box operator cannot (Sec. 5.5); the same idea extends to transient heat equations both through implicit time stepping and through a second-kind causal space-time boundary integral that supports direct arbitrary-time queries.

Limitations. The learned radial law is validated for simply-connected interior screened problems and in-range parameter interpolation; it does not transfer unchanged to exterior or multi-component systems. On

variable-coefficient volume problems, the one-shot learned model improves over FNO by only $\sim$1.3$\times$; the stronger result there is numerical preconditioning rather than learned operator accuracy (Appendix A). Near-touching gaps below 0.005 remain sensitive to precision and quadrature, and the 3D study uses coarse radial surfaces rather than general meshes. Near-linear $\mathcal{H}$-matrix scaling is demonstrated for low-frequency Laplace problems, while high-frequency Helmholtz requires a directional factorization (Sec. 5.4).

Practical target. When the exact kernel and an adequate classical BIE implementation are available, NBIO is not justified as a more accurate forward solver: its zero-correction limit is that solver, and the CAD experiment shows that learning can be mildly harmful. The intended uses are (i) a differentiable solver layer inside inverse, nonlinear, or time-dependent algorithms; (ii) correction of approximate physics; (iii) a common operator interface across geometry and discretization; and (iv) continuous off-boundary and QBX evaluation. Pure forward-solve acceleration over optimized BIE software is not established.

6.1 Future work and reproducibility

The next numerical steps are higher-order 3D surface quadrature, adaptive double-precision QBX for very small gaps, and directional fast methods for high-frequency Helmholtz. The next learning question is not another generic architecture search, but whether similarly constrained normal forms can identify genuinely unknown coefficient-dependent kernels. These extensions lie outside the scope of the present study.

Reproducibility. Neural NBIO training and evaluation use float32 unless otherwise stated; the reference BIE, interface/volume-source, and $\mathcal{H}$-matrix calculations use float64, and the continuous-time heat operator reports both precisions (Table 18). Ground truth is a manufactured (exact) solution where available; otherwise the reference discretization error is estimated by refinement and convergence checks. Trained-model results use multiple seeds where noted (e.g. the interior-Laplace ablation, wrong-kernel, variable-coefficient, CAD-baseline, and joint-3D studies). An anonymized supplementary archive accompanying this submission contains the NBIO layers (analytic kernel split, differentiable second-kind solve, QBX close evaluation), the $\mathcal{H}$-matrix solver, the interface/volume-source assembly, the baselines, and the geometry families, boundary-data distributions, train/test splits, evaluation points, seed lists, and per-experiment configurations. It includes compact raw/aggregate result files and the selected checkpoints for the radial normal form and its matched learned controls; licensed external CAD meshes, generated tensors, and other checkpoints are omitted for redistribution and size. The included scripts recreate the synthetic data and controlled experiments used for the core claims.

7 Conclusion

NBIO combines two complementary sources of accuracy. Its boundary-integral core retains the stability and discretization behavior of the classical solver, while the radial specialization identifies a missing screened-kernel law from solution data. The κr normal form improves accuracy by nearly two orders of magnitude over matched generic parameterizations, interpolates across a fully withheld parameter interval, and remains effective under limited and noisy supervision. A same-protocol Boundary Transformer remains far less accurate.

The present learned-transfer result applies to simply-connected interior screened problems; extending the learned approximation across distinct integral-equation regimes remains open. Within that scope, the paper establishes a practical design principle: retain exact singular and solve structure, and spend learned capacity on the smallest symmetry- and dimension-consistent representation of the missing physics.

References

Kendall E Atkinson. *The Numerical Solution of Integral Equations of the Second Kind*. Cambridge University Press, 1997.

Ran Bi, Jingrun Chen, and Weibing Deng. XI-DeepONet: An operator learning method for elliptic interface problems. *Journal of Computational Physics*, 538:114164, 2025. doi: 10.1016/j.jcp.2025.114164.

Nicolas Boullé, Seick Kim, Tianyi Shi, and Alex Townsend. Learning Green's functions associated with time-dependent partial differential equations. *Journal of Machine Learning Research*, 23(218):1–34, 2022.

Zhiwei Fang, Sifan Wang, and Paris Perdikaris. Learning only on boundaries: a physics-informed neural operator for solving parametric partial differential equations in complex geometries. *Neural Computation*, 36(3):475–498, 2024. doi: 10.1162/neco_a_01647.

Kyriakos C. Georgiou, Constantinos Siettos, and Athanasios N. Yannacopoulos. Learning contractive integral operators with fredholm integral neural operators. *arXiv preprint arXiv:2604.03034*, 2026.

Craig R Gin, Daniel E Shea, Steven L Brunton, and J Nathan Kutz. DeepGreen: deep learning of Green's functions for nonlinear boundary value problems. *Scientific Reports*, 11(1):21614, 2021.

Xindi Gong, Dingcheng Luo, Thomas O'Leary-Roseberry, Ruanui Nicholson, and Omar Ghattas. Shape derivative-informed neural operators with application to risk-averse shape optimization. *arXiv preprint arXiv:2603.03211*, 2026.

Leslie Greengard and Vladimir Rokhlin. A fast algorithm for particle simulations. *Journal of Computational Physics*, 73(2):325–348, 1987.

James Hipperson, Jonathan Hargreaves, and Trevor Cox. JAX-BEM: Gradient-based acoustic shape optimisation via a differentiable boundary element method. *arXiv preprint arXiv:2604.21431*, 2026.

George C Hsiao and Wolfgang L Wendland. *Boundary Integral Equations*, volume 164 of *Applied Mathematical Sciences*. Springer, 2008.

Muhammad Idrees Khan and Hua-Dong Yao. Quadrature-aware complex-linear neural operator for boundary-to-field prediction in resonant acoustics. *arXiv preprint arXiv:2607.04407*, 2026.

Andreas Klöckner, Alexander Barnett, Leslie Greengard, and Michael O'Neil. Quadrature by expansion: A new method for the evaluation of layer potentials. *Journal of Computational Physics*, 252:332–349, 2013.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to PDEs. *Journal of Machine Learning Research*, 24(89):1–97, 2023.

Rainer Kress. Boundary integral equations in time-harmonic acoustic scattering. *Mathematical and Computer Modelling*, 15(3–5):229–243, 1991. doi: 10.1016/0895-7177(91)90068-I.

Rainer Kress. *Linear Integral Equations*, volume 82 of *Applied Mathematical Sciences*. Springer, 3rd edition, 2014.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Graph kernel network for partial differential equations. *arXiv preprint arXiv:2003.03485*, 2020.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations (ICLR)*, 2021.

Zongyi Li, Daniel Zhengyu Huang, Burigede Liu, and Anima Anandkumar. Fourier neural operator with learned deformations for PDEs on general geometries. *Journal of Machine Learning Research*, 24(388): 1–26, 2023a.

Zongyi Li, Nikola Kovachki, Chris Choy, Boyi Li, Jean Kossaifi, Shourya Otta, Mohammad Amin Nabian, Maximilian Stadler, Christian Hundt, Kamyar Azizzadenesheli, and Animashree Anandkumar. Geometry-informed neural operator for large-scale 3D PDEs. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023b.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, 2021.

Lu Lu, Xuhui Meng, Shengze Cai, Zhiping Mao, Somdatta Goswami, Zhongqiang Zhang, and George Em Karniadakis. A comprehensive and fair comparison of two neural operators (with practical extensions) based on FAIR data. *Computer Methods in Applied Mechanics and Engineering*, 393:114778, 2022.

Bin Meng, Yutong Lu, and Ying Jiang. Solving partial differential equations in different domains by operator learning method based on boundary integral equations. *arXiv preprint arXiv:2406.02298*, 2024.

Sepehr Mousavi, Siddhartha Mishra, and Laura De Lorenzis. Imposing boundary conditions on neural operators via learned function extensions. *arXiv preprint arXiv:2602.04923*, 2026.

Pawan Negi, Maggie Cheng, Mahesh Krishnamurthy, Wenjun Ying, and Shuwang Li. Learning domain-independent Green's function for elliptic partial differential equations. *Computer Methods in Applied Mechanics and Engineering*, 421:116779, 2024. doi: 10.1016/j.cma.2024.116779.

Evert J Nyström. Über die praktische auflösung von integralgleichungen mit anwendungen auf randwertaufgaben. *Acta Mathematica*, 54:185–204, 1930.

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.

Sumanta Roy, Stephen T Castonguay, Pratanu Roy, and Michael D Shields. ϕ-DeepONet: A discontinuity capturing neural operator. *arXiv preprint arXiv:2604.08076*, 2026.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

Haixin Wang, Jiaxin Li, Anubhav Dwivedi, Kentaro Hara, and Tailin Wu. BENO: Boundary-embedded neural operators for elliptic PDEs. In *International Conference on Learning Representations (ICLR)*, 2024.

Haochen Wu, Heng Wu, and Benzhuo Lu. Operator learning on domain boundary through combining fundamental solution-based artificial data and boundary integral techniques. *arXiv preprint arXiv:2601.11222*, 2026a.

Sidi Wu, Aiqing Zhu, Yifa Tang, and Benzhuo Lu. Solving parametric elliptic interface problems via interfaced operator network. *Journal of Computational Physics*, 514:113217, 2024. doi: 10.1016/j.jcp.2024.113217.

Zhengyu Wu, Jun Liu, and Wei Wang. A boundary integral-based neural operator for mesh deformation. *arXiv preprint arXiv:2602.23703*, 2026b.

A Additional extensions and diagnostics

These structure-consistent extensions reinforce the core conclusions and are deferred here for focus; each is referenced from the main text.

A.1 3D surfaces

We extend NBIO to 3D surfaces (deformed spheres, approximate Laplace single layer + learned correction). Unlike the earlier interior-only correction, the final model is trained jointly on interior and exterior problems. Over three seeds it improves the bare structural solve in both regimes: interior rel-L_2 1.42e−2 → 2.10e−3 ± 4.3e−4 and exterior 1.63e−2 → 2.27e−3 ± 3.2e−4. It transfers zero-shot to stronger deformations (interior/exterior 2.63e−3±5.6e−4 / 3.08e−3±1.6e−4) and doubled surface resolution (4.14e−3±5.9e−4 / 3.92e−3±8.0e−4). A jointly trained DeepSet-3D remains near unit error (1.06 interior, 1.09 exterior; 3 seeds). This is a coarse radial-surface proof of concept with approximate area weights and self-panel quadrature, not yet a validated general-mesh 3D BEM or a real-CAD 3D operator (Fig. 10 right).

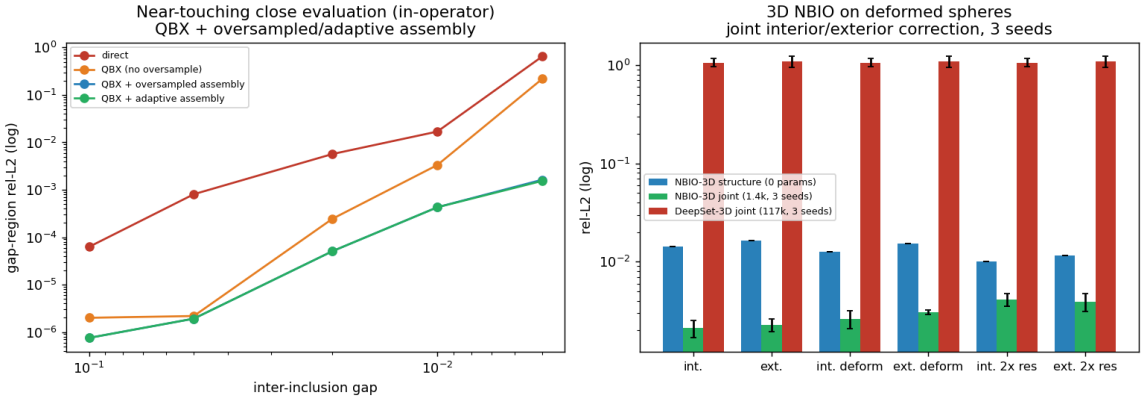

Figure 10: Left: near-touching close evaluation through the differentiable operator (QBX + oversampled/adaptive assembly). Right: jointly trained interior/exterior 3D NBIO on deformed spheres; bars show seed means and standard deviations for learned models.

A.2 Controlled runtime component audit

NBIO still assembles or applies a structured operator and solves a linear system per instance, so it does not have generic feed-forward surrogate latency. We therefore drop the earlier cross-hardware GPU/CPU ratio and report a *controlled* same-GPU component audit (NVIDIA TITAN X Pascal, float32, batch 8, 256 queries; 15 warm-up iterations followed by 80 per-repetition `cuda.synchronize`-bracketed timings, reported as median with interquartile range, per sample). At $n_b = 128$, structure-only assembly/solve/evaluation are 0.076/0.112/0.052 ms (component-summed total 0.24 ms, IQR 0.002); the learned split is 0.48/0.11/0.80 ms (component-summed total 1.39 ms, IQR 0.02). At $n_b = 256$ the component-summed totals are 0.90 and 3.82 ms. (Totals are the sum of the three independently-timed components, not a separately-timed end-to-end pass.) The learned pairwise correction, especially query evaluation, is the dominant overhead. These are controlled per-component GPU timings, not a claim of superiority over a tuned classical BIE or an amortized surrogate; the scalability claims in §5.4 remain algorithmic (iteration counts, rank, and asymptotic matvec cost).

For repeated use on one geometry we separately cache the LU factorization and, when query locations are fixed, the potential matrix. At $n_b = 128$, 256 queries, and batch 8, a new smooth full-prior geometry costs 1.24 ms/sample, whereas a new right-hand side with cached geometry/query operator costs 0.022 ms/sample; batched 32-right-hand-side solves cost 0.00120 ms per right-hand side and sample. Structure-only values are 0.191, 0.0222, and 0.00119 ms. These contracts are not interchangeable with a new-geometry

amortized surrogate, but they identify the practical repeated-load regime in which structured factorization reuse removes most online overhead. For the new screening-family study, the separately frozen dense new-geometry audit gives 3.88/6.54/10.80 ms per sample for exact/radial/generic-prior kernels at $n_b = 128$, one right-hand side, and 256 queries. This older component-oriented audit uses a different batch size and timing harness from the batch-size-one Transformer comparison in Sec. 4.2; the two timing sets are not directly comparable and are not combined into a speed ranking.

A.3 Transient problems: NBIO inside implicit time stepping

NBIO is not restricted to steady-state output maps when the time integrator reduces each step to an elliptic boundary-value problem. As a minimal parabolic demonstration, we solve the heat equation $u_t = \Delta u$ by implicit Euler. Each step requires $(-\Delta + \kappa^2)u^{n+1} = \kappa^2 u^n$ with $\kappa^2 = 1/\Delta t$, which we solve by a screened-FFT particular solution plus a homogeneous modified-Helmholtz boundary solve. A Yukawa-kernel NBIO trained only on fixed-κ modified-Helmholtz boundary data is dropped into the loop in place of the classical Yukawa BEM. With $\kappa = 4$, $\Delta t = 1/16$, and eight steps, NBIO matches the BEM-in-the-loop reference to 1.7×10^{-7} at final time; both have the same $\sim 1.65 \times 10^{-2}$ error to the exact heat-kernel manufactured solution, which is the shared implicit-Euler/FFT-particular floor rather than a boundary-solve error. This supports the intended scope: transient problems are accessible when a stable time discretization turns them into a sequence of elliptic solves, but NBIO is still not a generic recurrent time-series model. Because this experiment supplies the correct Yukawa kernel at each step, its accuracy is structural rather than evidence that a learned discrepancy model discovered heat dynamics.

solver in each implicit-Euler step	final rel-L_2 vs exact heat solution	final rel-L_2 vs BEM loop
Yukawa BEM reference	1.651e−2	—
Yukawa NBIO (structure-dom.)	**1.651e−2**	**1.665e−7**

Table 17: Heat equation by implicit Euler. NBIO tracks the classical screened-BEM solve inside the transient loop; the remaining exact-solution error is the shared time/particular-solution floor.

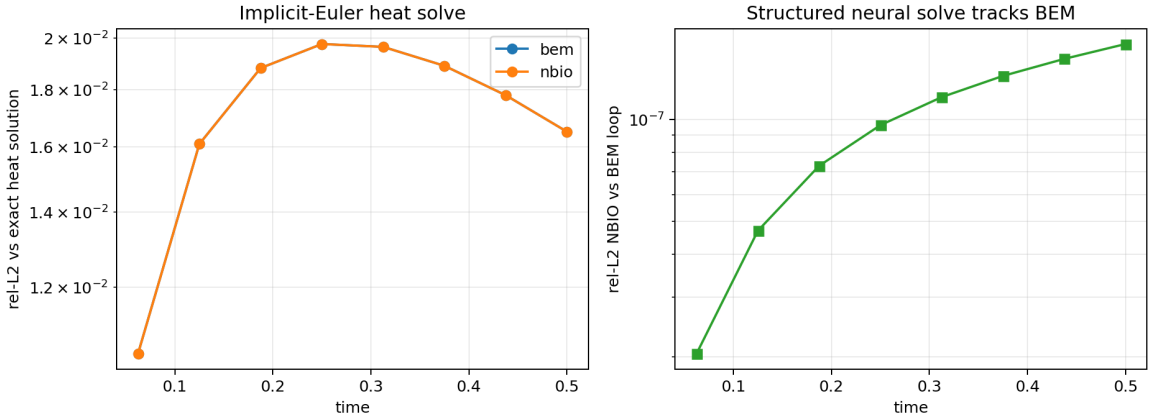

Figure 11: Parabolic heat demo. Left: BEM and NBIO have indistinguishable exact-solution error across implicit-Euler steps. Right: NBIO-vs-BEM loop error stays near the Yukawa NBIO solver floor.

Continuous-time parabolic NBIO. The implicit-Euler result above still reaches a target time by stepping. To test whether NBIO can support direct arbitrary-time queries, we also build a causal heat double-layer NBIO for zero-initial heat flow with time-dependent Dirichlet data: $u(x,t) = \int_0^t \int_\Gamma \partial_{n_y} H(x-y, t-s)\mu(y,s)\, ds_y\, ds$. The interior boundary trace is second-kind, $g(y,t) = -\frac{1}{2}\mu(y,t) +$ causal history, so the density is recovered by a block-lower-triangular Volterra solve and the solution can then be evaluated at arbitrary off-grid (x,t) queries. This fixes the first-kind caveat from a single-layer prototype: the second-kind double-layer version is stable in float32. On random off-grid time queries it reaches 2.1e−6 in float32 and

machine precision in float64, while a learned density map using the same analytic double-layer potential reaches 8.9e−2 and a black-box continuous-time DeepSet reaches 8.1e−1.

method	precision	arbitrary-time rel-L_2
parabolic double-layer NBIO	float32	**2.06e−6**
parabolic double-layer NBIO	float64	2.04e−16
learned continuous-time double-layer	float32	8.90e−2
continuous-time DeepSet	float32	8.10e−1

Table 18: One-shot arbitrary-time heat queries. The low error comes from the second-kind causal boundary-integral solve, not from a generic trajectory network.

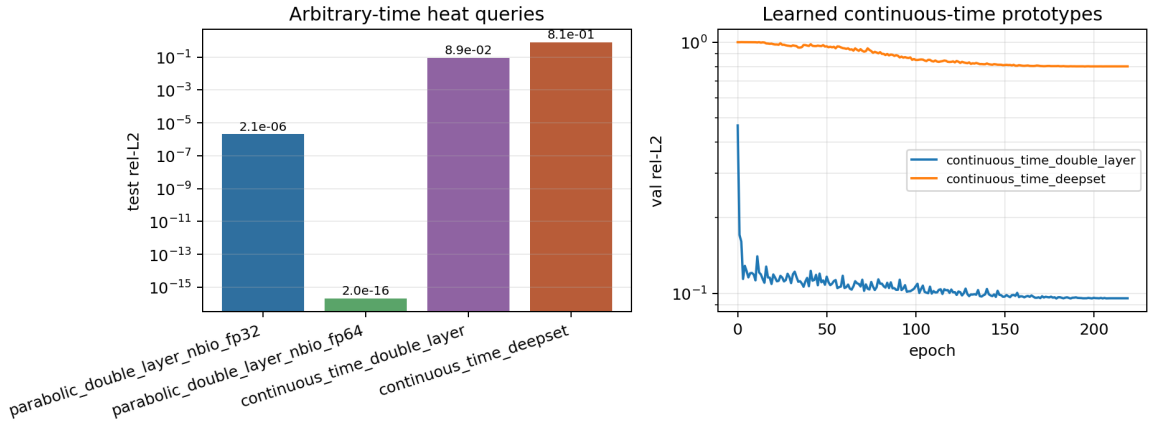

Figure 12: Continuous-time parabolic NBIO. Left: arbitrary-time query error. Right: learned continuous-time baselines improve with training but remain far above the second-kind causal solve.

Beyond Picard: Newton with a structured inner solve. The semilinear demo (Sec. 5.5) used Picard, which converges only while the fixed-point map contracts. Newton converges quadratically, and its step $J(u)w = -R(u)$ with $J = -\Delta + m^2 + 3\kappa u^2$ is a variable *reaction* (lower-order) operator – so the known Green's operator $(-\Delta + m^2)^{-1}$, shifted by the mean reaction, is still an exact preconditioner. On $-\Delta u + u + \kappa u^3 = f$, Newton with this structured inner solve reaches $\sim 10^{-15}$ in 5–13 iterations for all $\kappa \in [1, 100]$, where Picard converges only at $\kappa = 1$ and diverges beyond (Fig. 13); the inner GMRES count stays modest (7–88). The reaction *variation* (not its mean) is what still grows the inner count with κ – the same variable-coefficient-preconditioner frontier as above. The open case therefore remains a variable *principal* coefficient at high contrast, where the kernel itself must be learned.

Frequency limit (Helmholtz). The constant-rank and $O(1)$-iteration properties are *low-frequency* phenomena. On oscillatory Helmholtz ($\Delta u + k^2 u = 0$), measuring the same quantities versus the number of wavelengths $N = k\,\mathrm{diam}/2\pi$ across the domain (Fig. 16) shows the off-diagonal numerical rank growing $\sim$linearly with k ($9 \to 29$ over $N = 0 \to 32$), the best rank-4 block approximation degrading from $\sim 10^{-4}$ to 0.82, and GMRES iterations rising $7 \to 168$; a no-split separable learned operator trained on Helmholtz data correspondingly fails to reach 10^{-3} at any rank and worsens monotonically with k. This is the expected behaviour: Helmholtz Green's functions are high-rank in the far field. The classical remedy, the directional/butterfly fast multipole method, is demonstrated in Fig. 15: splitting the far field into $O(k)$ directional cones (half-width $\sim 1/(ka)$) and removing the source-side plane-wave phase keeps the per-cone numerical rank *bounded* ($\sim$7) while the raw block rank grows $7 \to 64$ over 2–128 wavelengths. Thus, the flat low-rank fast-solver story applies to non-oscillatory (Laplace-type) elliptic problems; high-wavenumber Helmholtz needs the known oscillatory kernel together with this k-dependent (directional) factorization.

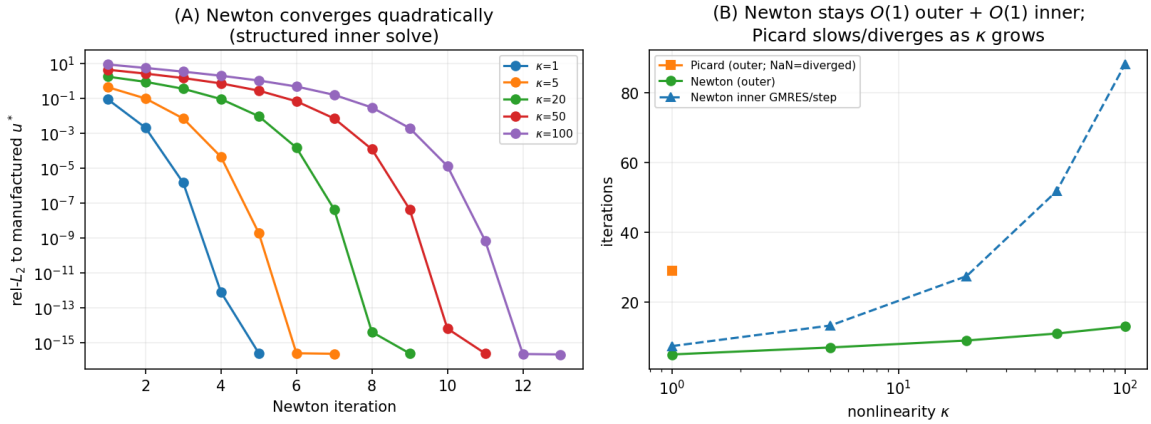

Figure 13: Newton for $-\Delta u + u + \kappa u^3 = f$ with a structured inner solve. (A) quadratic convergence to machine precision at every κ; (B) Newton stays $O(1)$ outer iterations (and $O(1)$-ish inner GMRES via the reaction-shifted known Green's preconditioner) while Picard slows and diverges as κ grows.

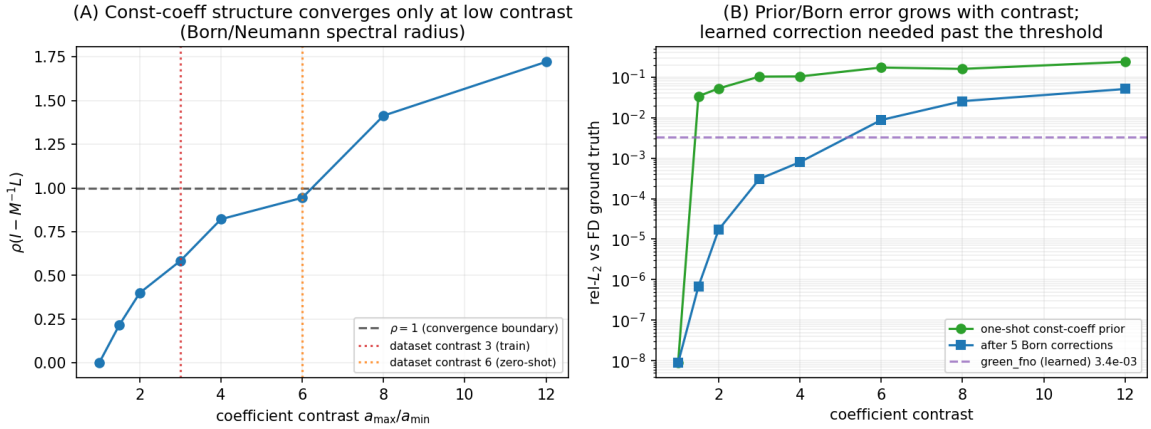

Figure 14: Variable coefficient $\nabla \cdot (a \nabla u) = f$. (A) the constant-coefficient preconditioner's iteration $T = I - M^{-1}L$ has spectral radius < 1 (converges) up to contrast $\sim$6; (B) where it converges, iterating the structured solver (Born/Neumann, no training) reaches the FD ground-truth floor, beating the learned green-FNO (3.4×10^{-3}). Past the divergence threshold a *one-shot* learned residual may help, but the same structured operator used as a GMRES *preconditioner* still reaches the ground-truth floor at every tested contrast *without* training (main text) – so learning is not required to solve the high-contrast problem, only to shortcut it in one shot.

B Parameter-family (κ) audit

Proposition 3 establishes that a model without κ as an input cannot represent the κ-family exactly. A frozen fixed-$\kappa = 4$ checkpoint was also evaluated with a κ-independent Fourier trace while varying κ. This is a joint boundary-load and operator-parameter stress test, not a pure parameter sweep: even at $\kappa = 4$, changing from the training point-source traces to Fourier traces raises split/no-split errors from 8.27e−3/2.85e−3 to 2.54e−2/4.78e−2. Error generally increases away from $\kappa = 4$, consistent with the non-identifiability established in Prop. 3. We do not turn the sampled 2× interval into a precise transfer radius, and we do not claim that the old aggregate numerically certifies the proposition's half-separation bound because the original diagnostic used different relative normalizations.

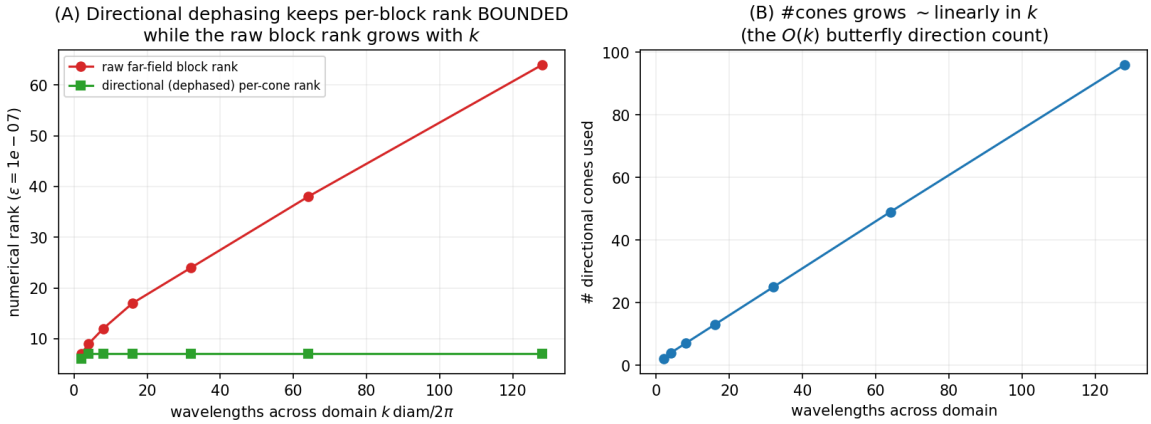

Figure 15: Helmholtz scalability fix. (A) the raw far-field block rank grows ∼linearly with the wavenumber, but directional dephasing within cones of half-width $\sim 1/(ka)$ keeps the per-cone rank bounded (∼7); (B) the number of cones grows ∼linearly in k – the $O(k)$ direction count of the directional/butterfly FMM, which restores $O(N \log N)$.

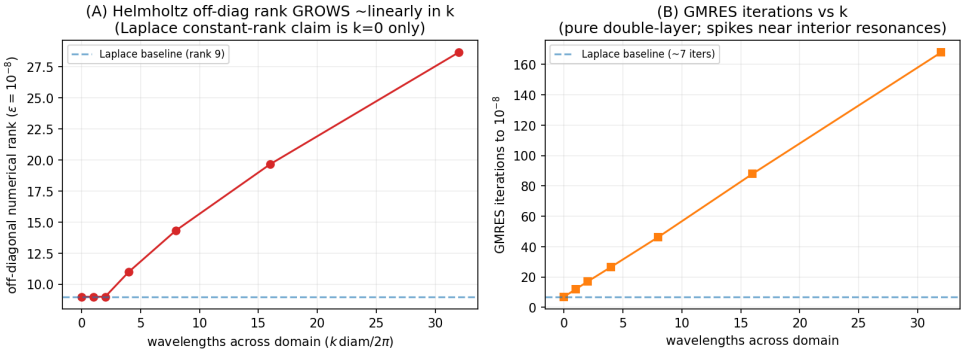

Figure 16: Frequency limit. For oscillatory Helmholtz the off-diagonal numerical rank (left) and the GMRES iteration count (right) both grow ∼linearly with the number of wavelengths across the domain, versus the flat Laplace ($k = 0$) baseline (dashed). The constant-rank / $O(1)$-iteration scalability properties are low-frequency phenomena.

Conditioned protocol. We therefore provide κ explicitly and train one correction over a stratified continuous $\kappa \in [0, 4]$ distribution: 2,400 training and 400 validation geometries at $n_b = 128$, including 10% mass at each endpoint. The feature is $\kappa/4$, and only the full-prior correction is multiplied by $(\kappa/4)^2$ so that model reduces exactly to the Laplace BIE at $\kappa = 0$; its analytic prior coefficient is fixed. The no-prior kernel is not multiplied by this envelope and must learn the complete kernel at every κ, including zero. Full-prior and no-prior models have equal capacity (9,089 parameters), use three independently generated training/validation datasets with paired initializations, train for 100 epochs with cosine decay, and select the best validation checkpoint. The untouched confirmation set has 96 new geometries and eight unseen in-range values $\{0.0625, 0.2, 0.6, 1.1, 1.6, 2.2, 3.0, 3.8\}$. Thus the reported confirmation comparison does not directly include the exact Laplace endpoint. The OOD aggregate likewise uses the strictly positive in-range values $\{0.2, 1.1, 2.2, 3.8\}$.

The regularity-controlled variant replaces normalized displacement/distance features by raw displacement, squared distance, e^{-r^2}, and smooth normal contractions. This removes the feature-level diagonal discontinuity while leaving width, depth, loss, and optimizer unchanged.

Alternative radial parameterizations. A 16-parameter positive monotone mixture reduced out-of-range degradation but was 5.88× worse than the radial MLP in range. A spatial cutoff of the Laplace prior introduced an annular transition that did not improve the family-level result. These comparisons motivated the unwindowed normal form in Eq. (1).

Endpoint and close-evaluation trade-offs. The exact screened multiplier has the local expansion

$$sK_1(s) = 1 + \frac{s^2}{2}\left(\log\frac{s}{2} + \gamma_E - \frac{1}{2}\right) + O(s^4 \log s),$$

so $m'_\star(0) = 0$. The original parameterization guarantees only $m_\theta(0) = 1$; its measured derivatives are $-0.0400, -0.0414, -0.0515$. For an off-surface approach to a source point, this leaves the learned remainder bounded but permits a direction-dependent pointwise limit. We tested two equal-parameter alternatives under fixed protocols. Endpoint subtraction, $m_\theta(s) = 1+s[f_\theta(s, \log(1+s)) - f_\theta(0,0)]$, enforces zero derivative but reaches 4.06e−4 versus 1.84e−4 for the frozen original model and 9.45e−5 for the exact BIE. A subsequent fresh holdout tested $m_\theta(s) = 1 + s^2 f_\theta(s, \log(1+s))$: it reaches 7.13e−4 versus 1.33e−4 original and 4.30e−5 exact. Both alternatives beat the matched full-prior and prior-free generic controls for every seed and remain well conditioned, but sacrifice bulk fidelity; the direct quadratic form also plateaus under resolution refinement. The $s^2 \log s$ term missing from its bounded smooth coefficient is a plausible explanation for that floor, not a separately identified cause. Endpoint subtraction improves pooled split-QBX error to 2.44e−4 (exact 7.24e−5), but the bulk-accuracy loss prevents it from replacing the primary model. We retain the accurate original form, use "bounded remainder" rather than "continuous regular remainder," and report these alternatives as local-regularity trade-offs.

We also tested the explicit asymptotic class $1 + s^2 \log(s) a_\theta(s) + s^2 b_\theta(s)$. It improves pooled split-QBX error from 8.74e−4 to 8.71e−5 and enforces the correct endpoint value and derivative, but its bulk error is 2.92e−4 versus 1.49e−4 for the selected model and it does not consistently improve from $n_b = 256$ to 512. We therefore retain the original radial model for the primary experiments and report the asymptotic variant as a close-evaluation trade-off.

Transfer beyond the training integral-equation regime. Without retraining, the interior radial checkpoints were applied to a single-obstacle exterior problem, a two-hole interior, and two- and four-obstacle exterior systems. The exact BIE remains between 8.4e−7 and 3.9e−6, whereas radial errors range from 0.76 to 10.1 despite finite solves and small residuals. Although the radial multiplier is geometry-independent by construction, the interior-trained checkpoint is not uniformly accurate over the separation and operator regimes induced by exterior or multi-component systems. The changed integral-equation conditioning then amplifies the kernel approximation error. We therefore restrict the learned-transfer claim to the simply-connected interior family used in training.

protocol slice	smooth full prior	smooth no prior	full-prior result
held-out confirmation, bulk ($n_b = 128$)	1.826e−3	2.816e−3	35.1% lower
held-out confirmation, near ($n_b = 256$)	4.459e−3	1.181e−2	62.2% lower
OOD point load, bulk ($n_b = 256$)	4.432e−3	4.173e−3	tied (CI crosses zero)
OOD point load, near ($n_b = 256$)	7.114e−3	1.651e−2	56.9% lower
OOD Fourier load, bulk ($n_b = 256$)	1.069e−2	2.718e−2	60.7% lower
OOD Fourier load, near ($n_b = 256$)	1.217e−2	3.746e−2	67.5% lower
out of range $\kappa = 5.25$, point bulk	6.066e−2	2.332e−2	160% higher

Table 19: Conditioned smooth-correction audit. Confirmation rows average three fresh training-data seeds over eight unseen in-range κ and 96 geometries. OOD rows additionally average four geometry families and 24 geometries per family; the Fourier trace is independent of κ. Paired geometry-cluster intervals exclude zero for the reported advantages except the explicitly tied point-bulk row. Correct Yukawa and fixed Laplace BIE controls are evaluated in every OOD cell.

At coarse $n_b = 128$ near-boundary point loads, the full prior is 20.3% worse than no-prior; after refinement to $n_b = 256$ it is 56.9% better. Thus retaining the analytic singular form is useful only when the quadrature

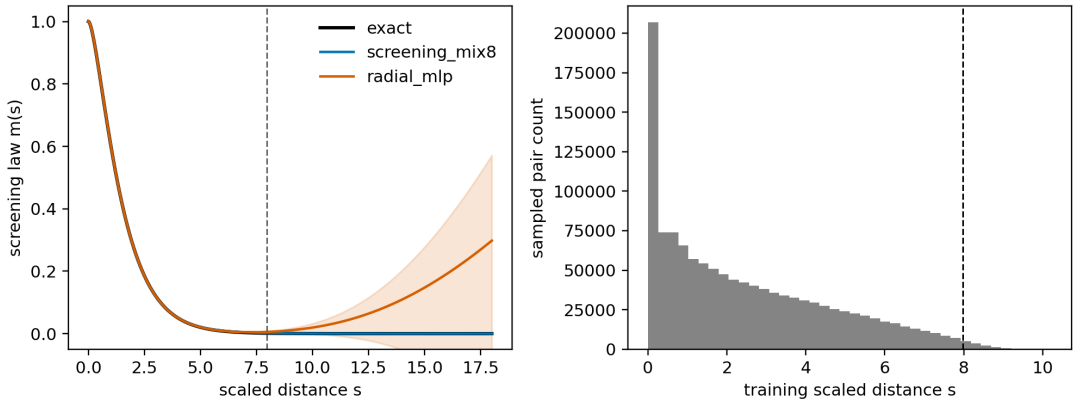

Figure 17: Post-hoc law diagnostic. Left: the three radial-MLP seeds (line and band) track the withheld exact multiplier $s\mathrm{K}_1(s)$ over densely supported separations; the monotone mixture is shown as an alternative parameterization. Right: empirical training-pair density. The dashed line is the 99th percentile of sampled s; divergence in the rare tail is why extrapolative law recovery is not claimed.

resolves it. Out-of-range bulk failure gives the second boundary: conditioning supports interpolation over the trained operator family, not unrestricted extrapolation. The smooth full-prior model is 3.8% lower than the legacy full-prior model in confirmation bulk (interval crosses zero) and 17.0% higher near the boundary, documenting a regularity/performance trade-off rather than a universal upgrade.

C Complete formulations and implementation details

Notation. Throughout this appendix, italic $K_0(x, y)$ denotes the Laplace double-layer boundary kernel, whereas upright $\mathrm{K}_0, \mathrm{K}_1$ denote the modified Bessel functions of the second kind; $\varkappa(y)$ denotes signed curvature and κ the screening parameter.

C.1 Interior, exterior, and Nyström discretization

For $G_0(x, y) = -(2\pi)^{-1} \log |x - y|$ and outward source normal n_y, our double-layer kernel is

$$K_0(x, y) = \partial_{n_y} G_0(x, y) = -\frac{1}{2\pi} \frac{(y - x) \cdot n_y}{|y - x|^2}.$$

On a smooth curve its same-panel diagonal limit under the periodic parameterization used here is $-\varkappa(y)/(4\pi)$, with signed curvature $\varkappa$. Uniform parameter nodes $t_j = 2\pi j/n_b$ receive arc weights $w_j = |x'(t_j)|2\pi/n_b$. We use

$$A_{ij} = \sigma\delta_{ij} + w_j\{(1 + \gamma)K_0(y_i, y_j) + K_\theta^{\mathrm{corr}}(y_i, y_j)\}, \qquad A\boldsymbol{\mu} = \mathbf{g},$$

with $\sigma = -1/2$ for the interior and $+1/2$ for the exterior. The query matrix uses the same weighted kernel without the jump term. The two-dimensional exterior double layer has a constant nullspace; the structure-only reference adds the rank-one weighted deflation $ww^\top/(\mathbf{1}^\top w)$ before solving. Log-singular single-layer self blocks in the transmission and Calderón formulations use the periodic Kress product rule; smooth cross blocks use the trapezoidal rule. QBX replaces only close evaluation of the gated analytic singular layer $(1 + \gamma)K_0$. The learned nonsingular correction K_θ^{corr} remains a direct weighted sum.

For modified Helmholtz, the correct reference uses $G_\kappa = (2\pi)^{-1}\mathrm{K}_0(\kappa|x - y|)$ and its normal derivative. The wrong-prior experiment deliberately keeps the Laplace K_0 in both assembly and evaluation while training the scalar gate γ and the nonsingular correction K_θ^{corr} at $\kappa = 4$. Thus the test measures discrepancy correction, not an exact Yukawa solver hidden in the model.

C.2 Correction network, initialization, and objective

The pairwise correction is a three-hidden-layer MLP of width 64 with GELU activations and one scalar output (9,026 parameters including the known-kernel gate in the standard fixed-κ model). The original input features are

$$(\widehat{y-x}, r, e^{-r}, \widehat{y-x}\cdot n_x, \widehat{y-x}\cdot n_y, n_x\cdot n_y, n_x),$$

where $r = |y-x|$ and $n_x = 0$ for an off-surface query. The output-layer weight is scaled by 0.05 and its bias is zero at initialization; the gate is $\gamma = \tanh(\zeta)$ with $\zeta = 0$. This is a near-classical, not exactly zero, initialization. The structure-only control explicitly zeros the correction and fixes $\gamma = 0$. We apply no reciprocity, symmetry, rotation-equivariance, or spectral-norm constraint. The relative features are translation invariant; arc weights make the integral insensitive to sampling density only to the accuracy of the chosen quadrature.

For the radial normal form, a three-hidden-layer width-32 GELU MLP receives $(s, \log(1+s))$, $s = \kappa r$, and outputs $m_\theta(s) = 1 + s f_\theta(s, \log(1+s))$ (2,241 parameters). The separate-input control replaces these two features by $(\kappa/4, r)$ and otherwise has the same architecture and parameter count. The matched generic control uses the smooth pairwise map below at width 30 (2,221 parameters). Radial training uses 2,400/400 geometries per seed, $n_b = 128$, three independent data/model/batch seeds, Adam at 10^{-3} with cosine decay, batch size 32, 7,500 updates, validation every 375 updates, and the lowest validation mean per-instance relative error. The frozen holdout has 32 new geometries in each of four families, two load families, five unseen in-range κ values, 256 queries, and $n_b \in \{64, 128, 256\}$. Checkpoints were hash-sealed before holdout evaluation.

For the conditioned family we use the regularity-controlled feature map

$$(y-x,\ r^2,\ e^{-r^2},\ (y-x)\cdot n_x, (y-x)\cdot n_y,\ n_x\cdot n_y,\ n_x,\ \kappa/4).$$

It has the same geometric feature count as the original map, plus the scalar parameter, and is smooth at $x = y$ for a smooth boundary and smooth normal field. Full-prior and no-prior conditioned models have equal capacity (9,089 parameters). The analytic prior coefficient is fixed in this experiment, and only its additive learned correction is multiplied by $(\kappa/4)^2$ so that it vanishes exactly at the Laplace endpoint. The no-prior kernel is ungated and learns the complete kernel, including at $\kappa = 0$. The conditioned protocol uses 2,400/400 training/validation geometries, $n_b = 128$, a stratified continuous $\kappa \in [0, 4]$ distribution, 100 epochs, and three independently generated data seeds with paired model initializations; the held-out confirmation and OOD sets are never used for checkpoint selection.

For samples i with query vectors u_i, training minimizes the instance-averaged relative error

$$\mathcal{L}(\theta) = \frac{1}{B}\sum_{i=1}^{B} \frac{\|\widehat{u}_i - u_i\|_2}{\max(\|u_i\|_2, 10^{-9})}.$$

The fixed-κ Laplace and Yukawa studies use 1,200/200/200 train/validation/test geometries, $n_b = 128$, 256 random query points per geometry, Adam with learning rate 10^{-3}, cosine decay, gradient clipping at 1, batch size 32, and 80 epochs. The best validation checkpoint is selected independently per seed. Geometry is a five-mode radial Fourier graph; boundary data are traces of six randomly weighted point sources outside the domain. There is no separate input/output normalization because the relative loss is scale invariant.

C.3 Transmission, prescribed jumps, and source split

For piecewise-constant conductivity with matrix coefficient a_0 and inclusion coefficients a_j, the indirect representation is $u = \sum_c S_{\Gamma_c}\phi_c$. The outer row enforces $\sum_c S_{0c}\phi_c = g$. On inclusion j, the flux condition gives

$$\left[-\frac{1}{2}\frac{a_0 + a_j}{a_0 - a_j} I + K^*_{jj}\right]\phi_j + \sum_{c\neq j} K^*_{jc}\phi_c = 0.$$

Each single-layer and adjoint-double-layer block may receive a learned nonsingular correction, although the reported exact-kernel structure-only rows set those corrections to zero.

For one interface with prescribed jumps $[u] = u_+ - u_- = \alpha$ and $[a\partial_n u] = a_+\partial_n u_+ - a_-\partial_n u_- = \beta$, we use the direct Calderón identity $\frac{1}{2}u + D[u] - S[\partial_n u] = 0$ in each region. The unknowns are outer flux q_0 and interface Cauchy data (u_+, q_+). We substitute $u_- = u_+ - \alpha$ and $q_- = (a_+q_+ - \beta)/a_-$ before assembling the three block rows. This uses single and double layers but no hypersingular block.

For $-\nabla\cdot(a\nabla u) = f$, the reported source experiment writes $u = N[f]/a_k + u_h$ in region k. The same periodic FFT Poisson potential N modifies the outer data and interface jump before the homogeneous Calderón solve, so its harmonic convention cancels on reconstruction. The implementation subtracts the box mean of f, uses a 1024^2 grid on $[-2.8, 2.8]^2$, periodic wrapping, and bicubic interpolation. Sources are compact Gaussians kept away from the box boundary. These assumptions are part of that experiment's contract; it is not a general free-space volume quadrature. One float64 scalar grid is 8 MiB; the complex FFT field is 16 MiB, before source, interpolation, and FFT workspace arrays. The boundary solve itself remains geometry-adaptive, but this particular-solution component is tied to the padded Cartesian box.

C.4 Assembly and evaluation pseudocode

For each geometry and boundary datum:

1. sample boundary coordinates, outward normals, curvature, and arc weights at the requested resolution;
2. evaluate the analytic self-kernel, replace its diagonal by the analytic limit, evaluate the pairwise correction, apply source weights, and add the interior/exterior jump;
3. solve the resulting geometry-dependent system for the density by a dense differentiable solve or the specified convergent fixed-point/Krylov method;
4. evaluate the known potential plus correction at arbitrary query coordinates, replacing the known close-evaluation term by QBX when its distance criterion is met.

Changing resolution repeats all four steps but does not update network parameters.

D Proof of Proposition 2

Step 1: continuum invertibility and stability. Write $A_\theta = A_0 + T_\theta = A_0(I + B)$ with $B = A_0^{-1}T_\theta : X \to X$, $X = C(\Gamma)$ with $\|\cdot\|_\infty$. Since $\|B\|_{X\to X} = \eta < 1$, the Neumann series $(I + B)^{-1} = \sum_{m\geq 0}(-B)^m$ converges in operator norm, so $A_\theta^{-1} = (I + B)^{-1}A_0^{-1}$ exists and

$$\|A_\theta^{-1}\|_{X\to X} \leq \frac{\|A_0^{-1}\|_{X\to X}}{1-\eta} =: C_\star. \tag{3}$$

Step 2: discretization and discrete stability. Let $t \mapsto x(t)$ be the 2π-periodic parameterization and $t_i = 2\pi i/n$, $i = 0, \ldots, n-1$, the collocation nodes ($h = 2\pi/n$). Define the sampling operator $\mathcal{R}_h : C(\Gamma) \to \ell_n^\infty$, $(\mathcal{R}_h\phi)_i = \phi(x(t_i))$, with the nodal norm $\|v\|_{h,\infty} = \max_i |v_i|$ ($\|\mathcal{R}_h\|_{X\to\ell_n^\infty} = 1$), and the periodic piecewise-linear extension $E_h : \ell_n^\infty \to C(\Gamma)$ satisfying $\mathcal{R}_hE_h = I$ and $\|E_h\|_{\ell_n^\infty\to X} = 1$. We distinguish two objects sharing the same kernel values. The *Nyström operator* $\mathcal{A}_{\theta,h} : X \to X$, $(\mathcal{A}_{\theta,h}\phi)(x) = \frac{1}{2}\operatorname{sgn}\phi(x) + \sum_j w_j\, K_\theta(x, x(t_j))\,\phi(x(t_j))$ (arc-length weights w_j, analytic diagonal/Kress correction for the log-singular self term), maps continuous functions to continuous functions. The *collocation matrix* $\mathbf{A}_{\theta,h} : \ell_n^\infty \to \ell_n^\infty$, $(\mathbf{A}_{\theta,h})_{ij} = \frac{1}{2}\operatorname{sgn}\delta_{ij} + w_jK_\theta(x(t_i), x(t_j))$, acts on nodal vectors; they are linked by $\mathcal{R}_h\mathcal{A}_{\theta,h} = \mathbf{A}_{\theta,h}\mathcal{R}_h$ on functions determined by their node values through the quadrature. The integral part of $\mathcal{A}_{\theta,h}$ is finite-rank; the family is *collectively compact* and converges pointwise to K_θ on X (standard for convergent Nyström quadrature of a continuous, weakly singular kernel). By the Anselone collectively-compact approximation theorem (Atkinson, 1997, Thm. 12.8), since $A_\theta = \frac{1}{2}\operatorname{sgn} I + K_\theta$ is invertible (Step 1), there is $h_0 > 0$ with $\mathcal{A}_{\theta,h}$ invertible on X for all $h < h_0$ and $\sup_{h<h_0}\|\mathcal{A}_{\theta,h}^{-1}\|_{X\to X} < \infty$. The matrix statement follows: for an arbitrary nodal right-hand side $b \in \ell_n^\infty$, set $g_h = E_hb \in X$ and $\phi_h = \mathcal{A}_{\theta,h}^{-1}g_h$; sampling and using $\mathcal{R}_h\mathcal{A}_{\theta,h} = \mathbf{A}_{\theta,h}\mathcal{R}_h$

gives $\mathbf{A}_{\theta,h}(\mathcal{R}_h\phi_h) = \mathcal{R}_h g_h = b$, so $\mathbf{A}_{\theta,h}^{-1}b = \mathcal{R}_h\mathcal{A}_{\theta,h}^{-1}E_h b$ and $\|\mathbf{A}_{\theta,h}^{-1}b\|_{h,\infty} \le \|\mathcal{A}_{\theta,h}^{-1}\|_{X\to X}\|b\|_{h,\infty}$ (both $\mathcal{R}_h$ and E_h have norm one). Hence eventual invertibility and the uniform bound transfer to the matrices,

$$\sup_{h<h_0} \|\mathbf{A}_{\theta,h}^{-1}\|_{\ell_n^\infty\to\ell_n^\infty} =: C_A < \infty. \tag{4}$$

This bound is in the nodal ℓ^∞ norm and does *not* equal an unweighted Euclidean matrix condition number.

Step 3: consistency defects. Define the operator and potential quadrature defects

$$\varepsilon_{A,h} := \big\|\mathcal{R}_h(A_\theta\mu_\star) - \mathbf{A}_{\theta,h}\mathcal{R}_h\mu_\star\big\|_{h,\infty}, \qquad \varepsilon_{P,h} := \big\|P_\theta\mu_\star - P_{\theta,h}\mathcal{R}_h\mu_\star\big\|_Y,$$

i.e. the errors made by replacing the exact boundary and potential integrals by their quadratures on the exact density $\mu_\star$.

Step 4: error decomposition. Let $\mu_\star = A_\star^{-1}g$ solve the target equation and $\mu_{\theta,h} \in \ell_n^\infty$ the discrete equation $\mathbf{A}_{\theta,h}\mu_{\theta,h} = \mathcal{R}_h g$; write $u_{\theta,h} = P_{\theta,h}\mu_{\theta,h}$. Splitting $\mathcal{R}_h\mu_\star - \mu_{\theta,h} = \mathbf{A}_{\theta,h}^{-1}\big[(\mathbf{A}_{\theta,h}\mathcal{R}_h\mu_\star - \mathcal{R}_h g)\big]$ and using $A_\star\mu_\star = g$ together with the model gap $A_\theta - A_\star = T_\theta - T_\star$ gives

$$\|\mathcal{R}_h\mu_\star - \mu_{\theta,h}\|_{h,\infty} \le C_A\big(\varepsilon_{A,h} + \|T_\theta - T_\star\|\,\|\mu_\star\|_X\big),$$

by (4). Note that in this line $\varepsilon_{A,h}$ is an *absolute* defect on $\mu_\star$ and is *not* multiplied by $\|\mu_\star\|_X$, whereas the model-gap term is. Mapping through the discrete potential $P_{\theta,h}$, adding and subtracting $P_\theta\mu_\star$, and using $\varepsilon_{P,h}$,

$$\|u_\star - u_{\theta,h}\|_Y \le \varepsilon_{P,h} + C_P C_A\,\varepsilon_{A,h} + \big(\|P_\theta - P_\star\|_{X\to Y} + C_P C_A\|T_\theta - T_\star\|_{X\to X}\big)\|\mu_\star\|_X, \tag{5}$$

where $C_P := \sup_{h<h_0}\|P_{\theta,h}\|_{\ell_n^\infty\to Y}$ is the uniform norm of the *discrete* potential-evaluation map on nodal densities. For queries at fixed separation from Γ (or through the stable QBX evaluation used for close queries) the quadrature weights are absolutely bounded and $C_P < \infty$; this boundedness is an assumption of the setting, stated explicitly. Writing $C_{\text{stab}} = \max(1, C_P C_A)$, which by (3)–(4) is finite, and bounding each coefficient by it gives the compact form $\|u_\star - u_{\theta,h}\|_Y \le C_{\text{stab}}\big[\varepsilon_{A,h} + \varepsilon_{P,h} + (\|T_\theta - T_\star\| + \|P_\theta - P_\star\|)\|\mu_\star\|_X\big]$ stated in Proposition 2; crucially the ε defects are not rescaled by $\|\mu_\star\|_X$.

Step 5: quadrature rates. For the periodic trapezoidal rule with Kress-type splitting of the logarithmic singularity, if the parameterization, data, and the *complete* kernels are C^p then $\varepsilon_{A,h}, \varepsilon_{P,h} = O(h^p)$; for C^∞ periodic integrands the rate is superalgebraic, and only for integrands admitting a bounded analytic extension to a strip is it geometric ($O(\rho^{-n})$) (Kress, 2014; Atkinson, 1997). For the original learned correction K_θ^{corr}, the MLP itself is C^∞ in its *feature* inputs (smooth activations), but the feature map contains the distance $r = |y - x|$ and the normalized displacement $\widehat{y - x}$, which are not smooth across the diagonal $x = y$; the *composite* boundary kernel is therefore not automatically C^p there, and the quadrature-rate conclusion is conditional on the resulting composite integrand. The conditioned regularity-controlled variant instead uses $y - x$, r^2, e^{-r^2}, and polynomial normal contractions. For a C^p boundary with a C^{p-1} normal field, this feature map removes the diagonal discontinuity and the composite correction has the corresponding finite regularity (up to the normal-field and activation regularity). Neither construction by itself guarantees an analytic extension, and we do not claim that a trained checkpoint inherits the classical superalgebraic or geometric rate. For the radial model, the boundary double-layer kernel K_0 has its analytic diagonal extension, while $m_\theta(\kappa r) - 1 = O(r)$. Hence $K_0\,[m_\theta(\kappa r) - 1]$ extends continuously to zero on the boundary diagonal. Because $r = |x - y|$ is not differentiable across that diagonal and the fitted models have $m_\theta'(0) \neq 0$, this extension is generally not differentiable there. For off-surface approaches, the same remainder is bounded but may have a direction-dependent pointwise limit, as audited in Appendix B. We therefore claim continuity of the boundary-operator correction, boundedness of the off-surface remainder, and no classical high-order quadrature rate for the learned component. Setting $T_\theta = T_\star = 0$ and $P_\theta = P_\star = P_0$ (the known-kernel case) reduces (5) to the classical Nyström convergence of the exact solver. □

E Comparison and reporting protocol

All core 2D Laplace/Yukawa models use the same 1,200/200/200 shape split and 256 query targets per instance. Per-instance relative errors are averaged; test geometry distributions are never used for checkpoint

selection. Training records the best validation checkpoint; the revision's learned-transfer and per-instance audits load that checkpoint. Legacy baseline aggregates evaluated the end of the same frozen schedule, without test-based selection. NBIO, no-split NBIO, DeepONet, and DeepSet use 80 epochs and the optimizer above. The Boundary Transformer uses four attention layers, width 128, four heads, 100 epochs, and batch size 24. GeoFNO uses a 40^2 latent grid, 16 Fourier modes, width 40, four layers, 120 epochs, and batch size 8; all three GeoFNO seeds use the same fixed splits. Parameter counts are reported in each result table. *Tuning budget.* Each model was trained with a *single* hyperparameter configuration (the settings above); we ran *no* per-model hyperparameter sweep, so the number of configurations attempted per baseline is one, and NBIO received no more configuration-search effort than any baseline. The only quantities selected were the standard optimizer settings above, held fixed across seeds. This is deliberately an engineering comparison, not an equal-number-of-trials tuning competition: a broader search could improve any of these baselines. All 2D models are small (NBIO 9k to GeoFNO 6.6M parameters) and train on 1,200 geometries for 80–120 epochs, i.e. minutes per model on one GPU (a few GPU-hours in total across all models and seeds). GINO was attempted but not scored because its canonical per-sample changing-geometry path did not complete the frozen training protocol.

Table 20 records the remaining same-protocol learned comparisons, including the exploratory extensions. An "update" is one optimizer step, not one full pass over a continuously generated distribution. Single-seed rows are disclosed rather than assigned artificial uncertainty.

comparison	train/val/test cases	queries	optimization and selection	learned seeds
2D Laplace	1200/200/200	256	Adam, 80 ep., bs 32 (2960 upd.); validation/frozen schedule	NBIO/no-split/GeoFNO 3; DNO/DS 2; BT 3
Yukawa $\kappa = 4$	1200/200/200	256	same core protocol; best validation checkpoint in transfer audit	split/no-split 3; DNO/DS 1
screening-family radial	2400/400/128 geom.	256	Adam, 7500 upd., bs 32; best validation, hash-sealed	radial/generic/BT 3
screening interval holdout	2400/400/96 geom.	256	same; train $[0, 1.5] \cup [2.5, 4]$, test $(1.5, 2.5)$	radial/generic controls 3
screening data efficiency	100–2400/400/128 geom.	32 or 256	same 7500 upd.; clean validation; fixed iid label noise	radial/generic 3
transmission	1000/200/200	256	Adam+cosine, 80 ep. (BT 100); frozen final epoch	split 1; BT 3; DS 2
interface jumps	800/150/60	256	Adam+cosine, 120 ep., bs 16 (6000 upd.); frozen final epoch	Transformer/DS 1
interface source+jumps	800/150/60	256	same interface protocol	Transformer/DS 1
ABC CAD	265/70/145	192 eval.	AdamW, at most 300 upd.; mesh-disjoint validation early stop	spatial-bias 3
Fusion CAD	210/45/95	192 eval.	AdamW, at most 300 upd.; mesh-disjoint validation early stop	NGF 3
joint 3D surface	generated/–/16 per slice	360	Adam, 140 generated-batch upd.; no checkpoint selection	NBIO/DS 3
continuous-time heat	192/48/64	160	AdamW+cosine, 220 ep., bs 8 (5280 upd.); best validation	layer/DS 1

Table 20: Training and selection contracts for same-protocol learned comparisons. DNO, DS, and BT abbreviate DeepONet, DeepSet, and Boundary Transformer. CAD test meshes are untouched until final evaluation; the previous test-subset early-stopping results are superseded and not used in this manuscript. Parameter counts appear beside results. Published context rows use neither these splits nor these ratios.

Data-efficiency status. The evaluation set in Table 8 had been used in earlier model development. Before this study, we fixed every training-data budget, query count, noise level, metric, and analysis, then selected checkpoints using only clean validation data and reported all conditions. We therefore treat the result as supporting evidence for unique-observation efficiency rather than an independent confirmation or a formal sample-complexity estimate.

The fixed-$\kappa = 4$ transfer audit uses 64 paired test instances per geometry family and query margin, three learned-model seeds, and $n_b \in \{64, 128, 256\}$. Geometry, manufactured sources, and query points are held fixed across resolution. Its machine-readable summary reports per-instance mean, median, p90, test-instance standard deviation, seed-mean standard deviation, and paired-bootstrap 95% intervals. This separates variation across test problems from variation across initialization seeds.

Tables 4, 14, and 15 explicitly separate identical-data comparisons from published context. The latter are not included in ratios or significance statements. NBIO reassembles and solves for every geometry, whereas learned baselines amortize training into a forward pass. Accuracy and transfer tables therefore compare output quality under different online computational contracts. Section 5.4 studies NBIO's algorithmic scaling; it does not establish lower total cost than an optimized classical BIE implementation.

margin	family	method	n_b=64	n_b=128	n_b=256
0.12	in-dist	correct Yukawa BIE (reference)	5.06e−4	1.35e−5	1.68e−6
		fixed Laplace BIE (wrong prior)	1.03e+0	1.03e+0	1.03e+0
		Laplace split + corr. (NBIO)	8.71e−3	8.62e−3	8.63e−3
		learned kernel (no split)	8.45e−3	3.14e−3	3.60e−3
	high-mode	correct Yukawa BIE (reference)	9.81e−4	2.15e−5	1.99e−6
		fixed Laplace BIE (wrong prior)	1.09e+0	1.09e+0	1.09e+0
		Laplace split + corr. (NBIO)	1.31e−2	1.31e−2	1.32e−2
		learned kernel (no split)	9.87e−3	4.80e−3	4.94e−3
	rough	correct Yukawa BIE (reference)	2.64e−3	6.47e−5	2.84e−6
		fixed Laplace BIE (wrong prior)	1.02e+0	1.02e+0	1.02e+0
		Laplace split + corr. (NBIO)	2.11e−2	2.08e−2	2.09e−2
		learned kernel (no split)	1.30e−2	6.94e−3	6.78e−3
	deform	correct Yukawa BIE (reference)	1.66e−3	7.21e−5	2.84e−6
		fixed Laplace BIE (wrong prior)	1.15e+0	1.15e+0	1.15e+0
		Laplace split + corr. (NBIO)	1.48e−2	1.41e−2	1.41e−2
		learned kernel (no split)	1.26e−2	4.77e−3	4.87e−3
0.02	in-dist	correct Yukawa BIE (reference)	9.63e−2	1.77e−2	9.89e−4
		fixed Laplace BIE (wrong prior)	7.21e−1	7.11e−1	7.10e−1
		Laplace split + corr. (NBIO)	1.28e−1	2.84e−2	1.65e−2
		learned kernel (no split)	3.93e−2	3.36e−2	3.27e−2
	high-mode	correct Yukawa BIE (reference)	9.39e−2	2.17e−2	1.60e−3
		fixed Laplace BIE (wrong prior)	7.10e−1	7.01e−1	7.00e−1
		Laplace split + corr. (NBIO)	1.26e−1	3.43e−2	1.87e−2
		learned kernel (no split)	4.16e−2	3.62e−2	3.52e−2
	rough	correct Yukawa BIE (reference)	1.35e−1	3.26e−2	3.74e−3
		fixed Laplace BIE (wrong prior)	6.57e−1	6.34e−1	6.32e−1
		Laplace split + corr. (NBIO)	1.91e−1	5.65e−2	2.92e−2
		learned kernel (no split)	4.72e−2	4.06e−2	3.95e−2
	deform	correct Yukawa BIE (reference)	1.22e−1	2.50e−2	2.38e−3
		fixed Laplace BIE (wrong prior)	7.45e−1	7.29e−1	7.28e−1
		Laplace split + corr. (NBIO)	1.63e−1	4.10e−2	2.01e−2
		learned kernel (no split)	4.28e−2	3.39e−2	3.23e−2

Table 21: Complete fixed-κ=4 wrong-kernel transfer audit: mean per-instance rel-L_2 over 64 paired geometries per cell, three model seeds, both query margins (0.12 far-field, 0.02 near-boundary), all four geometry families, and $n_b \in \{64, 128, 256\}$. The correct Yukawa BIE reference refines with n_b; the wrong fixed Laplace BIE fails (rel-$L_2 \sim 1$ at margin 0.12); the Laplace-split correction and the no-split learned kernel both repair the wrong prior but plateau above the reference error, and the no-split model is generally lower at the standard margin. This legacy diagnostic motivates the radial normal form in Table 6.