COORDINATE-SPECIFIC FACTOR DESIGN IN LOW-RANK PHYSICS-INFORMED SOLVERS

Anonymous authors
Paper under double-blind review

ABSTRACT

Low-rank physics-informed solvers use products of one-dimensional factors, but coordinates can have distinct physical structures. Suitable factors can improve accuracy, yet coupled changes in family, capacity, and assignment complicate attribution. We introduce DimSpec-LR, a coordinate-specific framework with a shared CP interaction and a unified interface for neural, dictionary-based, and hybrid factors. Parameter-matched interventions examine family selection, capacity allocation, and coordinate assignment under controlled training configurations. Experiments across four PDE families demonstrate that physically aligned factors can substantially improve accuracy under a fixed training budget. On a boundary-layer problem, the prescribed profile reduces relative error from 0.8546 to 1.59×10^{-3} compared with a generic homogeneous control. However, a richer homogeneous alternative achieves an even lower error of 4.04×10^{-4}, revealing that coordinate suitability does not necessarily require heterogeneous factor families. Additional experiments show that allocating spectral capacity to cover the required temporal frequencies improves accuracy and that target-frequency coverage alone does not determine trained performance. A separate source-specified CDR problem shows that fixed local factors remain effective without constructing an analytic target solution. Initialization and preconditioning controls clarify their contributions to complete-solver accuracy. These findings establish a controlled evaluation methodology for physically informed factor design and highlight the importance of distinguishing coordinate suitability, capacity allocation, and architectural heterogeneity.

1 INTRODUCTION

Physics-informed neural networks (PINNs) learn fields from differential equations and prescribed constraints rather than interior labels (Raissi et al., 2019). Their accuracy depends on the representation used for the coordinates: Fourier features and sinusoidal networks improve oscillatory fitting, while localized bases express sharp spatial variation (Tancik et al., 2020; Sitzmann et al., 2020; Daubechies, 1992).

Separable PINNs represent solutions as sums of products of one-dimensional factors, enabling efficient axiswise evaluation and differentiation on tensor grids (Cho et al., 2023). The interaction determines how factors are combined, while their representations determine which directional structures can be expressed efficiently (Vemuri et al., 2024; 2026). A sharp boundary layer and a smooth cross-stream oscillation may call for different representations, motivating coordinate-specific factor design. Yet changing a representation can also change its capacity and coordinate assignment, making the source of an accuracy gain difficult to identify.

This raises our central question: under a fixed low-rank interaction and given physical priors, how do factor-family selection, capacity allocation, and coordinate assignment affect trained solution accuracy? DimSpec-LR combines a shared CP interaction with a unified interface for neural, dictionary-based, and hybrid factors. Decoupling representation from interaction allows each coordinate's family and capacity to be specified without changing the low-rank formulation. Parameter-matched family replacement, capacity routing, and fixed-inventory permutation provide complementary tests; rich homogeneous controls test whether gains require heterogeneous families. The primary studies fix $R = 16$ and hold expert-designed profiles constant across seeds.

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

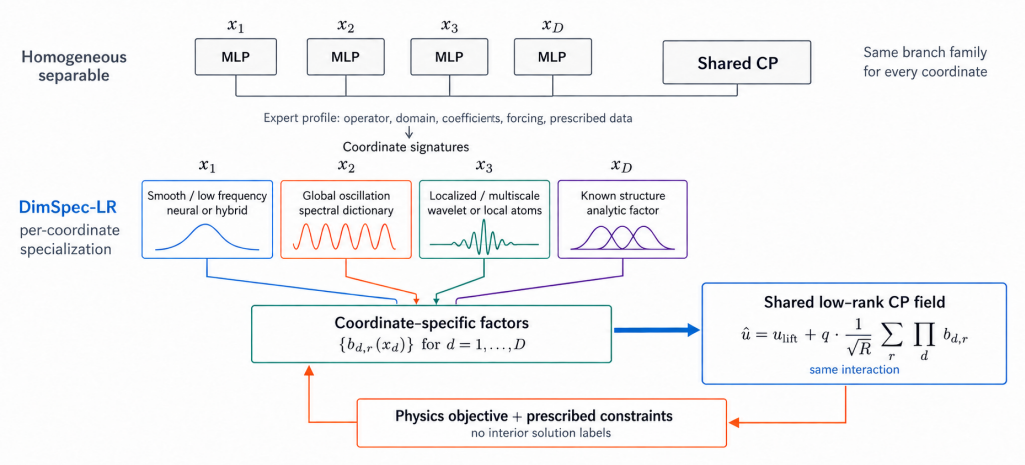


Figure 1: Factor design under a shared low-rank interaction. The illustrated heterogeneous profile is one option within DimSpec-LR; homogeneous families with equal or unequal capacities use the same interface. Profiles are expert-designed using the PDE specification, forcing, and prescribed data; training uses the physics objective and constraints.

The central finding is that *coordinate suitability does not require factor-family heterogeneity*. On A4 and transport case S3, permuting a fixed inventory harms accuracy, yet rich or capacity-routed homogeneous factors match or exceed the prescribed mixed profiles. On K3, covering the required temporal frequencies improves accuracy without requiring a particular capacity ratio.

Four PDE families provide complete-solver context; a separate source-specified CDR problem tests fixed profiles without an analytic target. This within-family test does not establish automatic selection or broad generalization. A restricted neutron scalar-flux application is reported separately.

The work makes three contributions:

- **A controlled evaluation protocol for factor-design attribution.** Parameter-matched family replacement, capacity routing, and fixed-inventory permutation, together with rich homogeneous controls, test competing explanations involving coordinate suitability, capacity, and heterogeneity.
- **A unified framework for controlled factor design.** DimSpec-LR supports neural, dictionary-based, and hybrid coordinate factors, allowing families, capacities, and assignments to vary independently of the shared CP interaction.
- **Empirical findings on representation suitability and its limits.** Assignment can affect accuracy without requiring heterogeneity; frequency-aware allocation improves finite-budget performance; target-mode coverage alone does not determine trained accuracy. Complete-solver comparisons, a source-specified CDR test, and targeted attribution controls complement these findings.

2 Related Work

Physics-informed coordinate models. PINNs encode equations and constraints in the optimization objective (Raissi et al., 2019). Fourier features and sinusoidal networks address spectral bias (Tancik et al., 2020; Sitzmann et al., 2020); PirateNet changes the trainable depth of a physics-informed network (Wang et al., 2024); and random-feature methods freeze nonlinear features before fitting output coefficients (Chen et al., 2022). These approaches motivate complementary representation choices. DimSpec-LR brings such choices to individual coordinate factors within a common low-rank formulation.

Functional tensor decompositions. SPINN combines coordinate networks through a low-rank outer product (Cho et al., 2023). CP-PINN and later functional tensor-decomposition work formalize

related CP, Tucker, and tensor-train views and compare ranks and activation backends (Vemuri et al., 2024; 2026). Factor-function selection is therefore an established design freedom. DimSpec-LR builds on it by placing neural, dictionary-based, and hybrid factors within a common CP formulation and examining how family selection, capacity allocation, and coordinate assignment affect trained accuracy. Capacity-routed controls and rich homogeneous alternatives test competing explanations for observed gains, rather than attributing them to architectural heterogeneity alone.

Operator learning and structured bases. FNO and DeepONet learn mappings between functions (Li et al., 2021; Lu et al., 2021); SepONet uses separate coordinate trunks in a physics-informed operator setting (Yu et al., 2024). DimSpec-LR instead optimizes one field per target PDE. Structured spectral, localized, and analytic bases enter through a common factor interface rather than as a new basis construction or an operator-learning claim.

3 Factor-Design Framework

3.1 Problem setting

Consider a field $u : \Omega \to \mathbb{R}$ on a product domain $\Omega = \Omega_1 \times \cdots \times \Omega_D$, governed by

$$\mathcal{N}_\lambda[u](\boldsymbol{x}) = f_\lambda(\boldsymbol{x}), \qquad \mathcal{B}_\lambda[u](\boldsymbol{x}) = g_\lambda(\boldsymbol{x}), \qquad \boldsymbol{x} = (x_1, \ldots, x_D). \tag{1}$$

Coordinates may have different physical meanings and approximation needs: layers, oscillations, smooth variation, or analytic constraints. Choosing one-dimensional representations independently of the interaction allows coordinate-specific design, including homogeneous factors where one family suits several coordinates.

3.2 Shared CP interaction

When a hard constraint construction is available, DimSpec-LR writes

$$\widehat{u}(\boldsymbol{x}) = u_{\text{lift}}(\boldsymbol{x}) + q(\boldsymbol{x}) \frac{1}{\sqrt{R}} \sum_{r=1}^{R} \prod_{d=1}^{D} b_{d,r}(x_d; \theta_d). \tag{2}$$

The lift satisfies nonhomogeneous data; q vanishes on hard-constrained boundaries. Soft-only constraints use $u_{\text{lift}} = 0$, $q = 1$. Main profiles and matched interventions fix $R = 16$; sensitivity checks vary it.

CP keeps interaction fixed without Tucker's learned cross-factor core or tensor trains' sequential cores; comparing interactions is outside our scope. We vary factors through

$$b_{d,r}(x_d) = \boldsymbol{\Phi}_d(x_d)^\top \boldsymbol{c}_{d,r} + \eta_d n_{d,r}(x_d), \tag{3}$$

where $\boldsymbol{\Phi}_d$ is a dictionary with trainable coefficients $\boldsymbol{c}_{d,r}$. Each coordinate network $n_d(x_d) \in \mathbb{R}^R$ has R outputs, with $n_{d,r}$ its r-th channel. Dictionary-only factors use $\eta_d = 0$; neural-only factors omit the dictionary and set $\eta_d = 1$. Corrections use $\eta_d = \text{softplus}(\gamma_d)$ with trainable scalar γ_d (Section A.2); fixed coefficients permit analytic factors.

The interface separates three choices from the interaction: *family* specifies the representation, *capacity allocation* distributes atoms or neural parameters, and *coordinate assignment* maps a fixed factor inventory to axes. Both homogeneous and heterogeneous profiles are supported. Interventions (Section 4) are targeted comparisons, not a full factorial decomposition. Capacity matching concerns trainable parameter counts, not identical function spaces.

3.3 Expert-designed factor profiles

Expert-designed profiles let us examine how physical information can be encoded in coordinate factors and tested through controlled interventions. Table 1 groups the choices by directional signature.

Profiles are benchmark-aware and fixed across seeds. Their physical cues come from the operator, domain, coefficients, forcing, and prescribed boundary/initial data, including an initial time derivative. Interior reference values do not enter training, checkpoint selection, or initialization.

Table 1: Coordinate signatures and corresponding branch choices. The symbols are defined exactly in Section A.2.

Signature	Branch family	Suitable behavior
Smooth / low frequency	compact MLP, SIREN, hybrid H_m	slowly varying spatial or temporal factors
Global oscillation	sine S_m, Fourier features	periodic or narrow-band coordinates
Localized / multiscale	boundary atoms B_m, sparse DB4	layers, resonances, and local coefficient changes
Known structure	analytic factor, collision C_m	endpoint, symmetry, or angular constraints
Imperfect prior	$P(B; w, d)$	dictionary structure with an off-dictionary correction

Table 10 audits the distinction between physical cues and expert-defined modeling choices. Bandwidths and correction widths are not uniquely determined by the PDE. This setting evaluates how specified priors affect trained accuracy, including with fixed CDR profiles (Section E.1); it does not define an automatic selection rule.

3.4 Why factor choice matters

The following approximation argument separates interaction complexity from one-dimensional representation error. Suppose

$$u = u_{\text{lift}} + qR^{-1/2} \sum_{r=1}^{R} \prod_{d=1}^{D} f_{d,r} + e_R, \qquad \|e_R\|_{L^2(\Omega)} \le \delta_R, \tag{4}$$

and let $\|f_{d,r} - b_{d,r}\|_2 \le \varepsilon_{d,r}$ with $\max(\|f_{d,r}\|_2, \|b_{d,r}\|_2) \le M_{d,r}$. Then

$$\|u - \widehat{u}\|_2 \le \delta_R + \frac{\|q\|_\infty}{\sqrt{R}} \sum_{r=1}^{R} \sum_{d=1}^{D} \varepsilon_{d,r} \prod_{j \ne d} M_{j,r}. \tag{5}$$

Equation (5) separates interaction approximation error from individual coordinate-factor contributions (proof in Section C). With the CP interaction and norm bounds fixed, it motivates reducing directional approximation errors through suitable families and capacity allocation. The argument applies to homogeneous and heterogeneous representations alike: it motivates coordinate suitability, not heterogeneity itself. Products of norm bounds can grow with dimension, and individual factor norms depend on the CP scaling convention. The bound is therefore not a dimension-uniform accuracy guarantee. It also does not characterize optimization dynamics or strong-form residual errors; controlled finite-budget experiments test how factor choices translate into trained accuracy.

3.5 Physics-only training and cost

Let $\mathbf{r}_g(\widehat{u})$ be one residual block evaluated on collocation set $\mathcal{C}_g$. The objectives have the general form

$$\mathcal{L}_{\text{phys}} = \sum_{g \in \mathcal{G}} \frac{w_g}{|\mathcal{C}_g|} \|T_g \mathbf{r}_g(\widehat{u})\|_2^2 + \sum_j \gamma_j \mathcal{L}_j^{\text{soft}}, \tag{6}$$

where T_g weights or preconditions a residual block; solution-dependent normalizers use stopped gradients. The neutron objective is specified in Equation (54). Soft terms enforce constraints outside the hard ansatz. Matched arms share loss, constraints, sampling, and update budget; complete solvers retain casewise ansatzes and loss implementations. No interior reference values enter the loss. Equal updates are not equal compute: wall-clock time also depends on factor cost, validation, and hardware load.

For n_d samples along coordinate d, storing all factor evaluations costs $O(R \sum_d n_d)$. A dictionary with m_d atoms requires $O(n_d m_d R)$ operations to form its factors; a neural factor incurs its layer cost. This saving concerns axiswise evaluation and differentiation: explicitly assembling the dense tensor grid still costs $O(R \prod_d n_d)$ arithmetic and requires $O(\prod_d n_d)$ output storage. Dense physical operators can still dominate total runtime, especially for integral transport terms. Smaller factors therefore need not produce a faster complete solver.

Table 2: Four PDE solver families, 16 cases. The additional N1–N4 restricted scalar-flux applications are specified separately in Section B.5. Exact equations and references are given in Section B.

Family	Four case roles	Reference construction
Helmholtz	H1: curved wavefront, coupled coefficient; H2: off-dictionary frequencies; H3: unequal axis frequencies; H4: four spectral pairs	H1 numerical; H2–H4 manufactured
Anisotropic CDR	A1: curved transported plume; A2: unequal directional scales; A3: globally rotated anisotropy; A4: sharp outflow layer	A1 numerical; A2–A4 manufactured
Stationary transport	S1: high-order angular kernel; S2: isotropic scattering; S3: forward anisotropy; S4: three-material interface	Independent discrete-ordinates reference
Klein–Gordon	K1: smooth 2D space–time; K2: additional spatial axis; K3: fast temporal oscillation; K4: multimode unequal scales	Manufactured

4 Experiments

4.1 Benchmark design and protocol

Matched interventions test explanations of factor-design gains under a shared CP interaction; complete-solver comparisons assess practical performance. Fixed-profile, scale, rank, and budget checks examine applicability.

The main PDE suite has 16 cases: four each from Helmholtz, anisotropic convection–diffusion–reaction (CDR), stationary transport, and Klein–Gordon families (Table 2). A source-specified CDR test evaluates fixed profiles separately (Table 5). N1–N4 are separate restricted scalar-flux applications, not full angular-transport solver tests (Section B.5 and Table 14).

Main DimSpec-LR profiles and matched controls use $R = 16$. Gradient-trained H/A/S/K models use 3,000 optimizer updates, our primary finite-budget setting. We report relative L_2 field error for H/A/K and scalar-flux error for S/N, alongside measured training time and trainable parameter count. Checkpoints minimize held-out physics validation; monitored reference errors never select them. Profiles are expert-designed, fixed across seeds, and benchmark-aware (Table 10). Reported training times condition on the specified profiles and exclude configuration-selection costs.

The declared baselines are Vanilla PINN, MS Fourier PINN, SPINN, PirateNet, and RFM (Tables 11 and 12). SPINN is the closest homogeneous low-rank neural baseline. Neural methods receive the same casewise update count; RFM uses its prescribed linear physics fit. Complete-solver comparisons retain each method's constraints, preconditioning, and initialization. For example, A3 uses an operator-based projected initialization not shared by the baselines, and A4 constraint implementations differ (Section A.3). These comparisons therefore measure complete configurations rather than isolated representation effects.

Three seeds are used for the matched studies and the four representatives; the separate scale and rank sensitivities use seed 42. The remaining cases form a single-seed coverage audit, not a population win-rate estimate. Casewise training configurations and replication settings are specified in Section A.4; H2/A2/A3/K2 replication measures profile stability against single-seed comparators. Solver comparisons use case-specific hyperparameters, not a common configuration-selection budget.

4.2 Factor suitability and coordinate assignment

We first test factor suitability, then coordinate correspondence, and finally whether the gains require heterogeneous families (Table 3). A4 pairs a thin outflow layer with a smooth cross-stream oscillation; S3 couples space and angle through forward-anisotropic scattering. Homogeneous, capacity-routed, permuted, and prescribed profiles share each case's loss, collocation stream, optimizer, and 3,000-update budget, with parameter counts within 0.5% of the prescribed profile. S3 uses learning rate 5×10^{-4} and depth-2 neural corrections; its results are separate from the main S3 profile trained at 0.002.

Table 3: Factor family, assignment, and homogeneous alternatives. Mean ± sample standard deviation over seeds 42, 2026, 3407; all arms use $R = 16$ and 3,000 updates. A4 measures field error; S3 measures scalar-flux error. All S3 branches have depth-2 corrections, width 48 except where shown. The lowest mean in each study (bold) is attained by a homogeneous control.

Case	Control	Profile	Parameters	Relative L_2
A4	Homogeneous, uniform	H_{52}/H_{52}	6,784	0.8546 ± 0.0016
A4	Homogeneous, routed	H_{72}/H_{31}	6,720	0.8484 ± 0.0090
A4	Permuted	H_{32}/B_{64}	6,752	0.8519 ± 0.0033
A4	Prescribed	B_{64}/H_{32}	6,752	$1.59 \times 10^{-3} \pm 4.15 \times 10^{-4}$
A4	Rich family on both axes	B_{44}/B_{44}	6,720	$\mathbf{4.04 \times 10^{-4} \pm 1.77 \times 10^{-4}}$
S3	Homogeneous, uniform	H_{36}/H_{36}	11,202	0.1390 ± 0.0157
S3	Homogeneous, routed	H_{64}/H_7	11,138	$\mathbf{0.0394 \pm 0.0039}$
S3	Permuted	C_{32}/H_{64}	11,154	0.2931 ± 0.0206
S3	Prescribed	H_{64}/C_{32}	11,154	0.0448 ± 0.0015
S3	Collision, uniform widths	C_{32}/C_{32} (62/62)	11,134	0.0467 ± 0.0015
S3	Collision, routed widths	C_{32}/C_{32} (74/48)	11,188	$0.0445 \pm 5.73 \times 10^{-4}$

A: Physically aligned factors improve finite-budget accuracy. A4 is a controlled setting whose PDE coefficient specifies the layer scale. The B dictionary contains the target layer atom and H the cross-stream sine component, making the target representable by a single product up to normalization (Equations (11) and (33)). Under matched training, prescribed B_{64}/H_{32} reduces mean error from uniform H_{52}/H_{52}'s 0.8546 to 1.59×10^{-3}. Routing more H-family capacity to the layer axis yields 0.8484: redistributing the tested generic factors does not substitute for the aligned localized representation at this budget. This demonstrates the benefit of an explicitly specified physical prior, rather than automatic adaptation to an unknown layer shape.

B: Assignment matters within a fixed inventory. Are suitable components sufficient regardless of placement? Permuting B_{64}/H_{32} to H_{32}/B_{64} raises A4 error to 0.852, despite retaining the same inventory and parameter count. On S3, prescribed H_{64}/C_{32} reaches 0.0448 scalar-flux error, compared with 0.293 for its permutation. Thus, the same inventory can perform very differently depending on its coordinate correspondence. The S3 finding concerns scalar flux, not the ordering of full angular-field or higher-moment errors.

C: Suitable factors need not be heterogeneous. The final comparison shows that coordinate suitability and heterogeneity are distinct properties. A4's rich homogeneous B_{44}/B_{44} control reaches 4.04×10^{-4}, outperforming prescribed B/H at essentially the same parameter count. Both axes receive hybrid and boundary atoms at reduced bandwidth. On S3, capacity-routed H/H reaches 0.0394 and C/C variants reach 0.0445–0.0467, versus 0.0448 for H/C. A4 B/B and S3 routed H/H outperform their prescribed profiles at all three seeds (Table 17); these are counterexamples to heterogeneity being necessary, not statistical equivalence tests.

Both A4 profiles contain the exact rank-one target and recover the layer (Figure 2, appendix), yet B/B trains to lower error. Target inclusion alone cannot explain this difference. Together, the three comparisons establish the benefit of suitable representations, the importance of their placement, and the viability of rich homogeneous alternatives.

4.3 Matched spectral factors and capacity allocation

H4 asks whether target-mode coverage determines trained accuracy; K3 asks where spectral capacity should be placed. Each case fixes rank, sampling, loss, constraints, and optimizer (Section E). H4 uses a common endpoint projection and approximately 2,240 parameters; K3 uses exactly 2,304. These protocols differ from the main solver configurations.

H4: spectral coverage is not sufficient. Every tested dictionary contains all H4 target modes, yet under the shared strong residual, projected H_{17} reaches mean error 0.0200, compared with 0.0947 for sine/cosine and 0.5478 for S_{70}. Equal target-mode coverage and parameter counts thus yield different trained accuracies; the approximately matched tanh control reaches 1.80.

Table 4: Independent matched interventions, 3,000 updates and three seeds. Errors are means ± sample standard deviations; time includes validation and monitoring on shared hardware. Every H4 branch uses endpoint projection Π, with $\Pi S_{70} = S_{70}$. These protocols differ from the main solver profiles.

Case	Profile	Parameters	Relative $L_2 \downarrow$	Time (s)
H4	S_{70}/S_{70}	2,240	$0.5478 \pm 2.02 \times 10^{-4}$	125.5
H4	$\Pi H_{17}/\Pi H_{17}$	2,240	$\mathbf{0.0200 \pm 6.85 \times 10^{-3}}$	165.8
H4	$\Pi SC_{35}/\Pi SC_{35}$	2,240	0.0947 ± 0.0224	144.4
H4	Projected tanh/tanh	2,232	1.7987 ± 0.4561	143.4
K3	$H_4/H_4/H_{25}$	2,304	$\mathbf{0.0150 \pm 3.73 \times 10^{-3}}$	391.4
K3	$H_{11}/H_{11}/H_{11}$	2,304	0.0838 ± 0.0133	395.9
K3	$H_{25}/H_4/H_4$	2,304	$1.0555 \pm 6.36 \times 10^{-3}$	410.7
K3	$H_4/H_{25}/H_4$	2,304	1.0644 ± 0.0111	391.2

The main S_{16}/S_{16} solver uses a smaller bandwidth and spectral preconditioning. Its substantially lower error therefore cannot be attributed to the sine factor family alone, and its results are not directly comparable to those of the matched intervention study.

A paired column-RMS control preserves initial fields and uses the same CPU (Table 18). Normalization leaves H_{17} better at every seed: column amplitudes cannot fully explain the gap. Target representability and finite-budget trainability are distinct. Derivative scales, Gram spectra, and non-target modes still differ, so no unique mechanism is identified.

K3: frequency-aware capacity allocation. The time-routed $H_4/H_4/H_{25}$ profile reaches mean error 0.0150, versus 0.0838 for uniform $H_{11}/H_{11}/H_{11}$. Moving the large factor to either spatial axis raises error to approximately 1.06. The time-routed profile is better at every paired seed, with the H family unchanged. It supplies $\sin(20t)$ and $\cos(20t)$ through H_{25} on the time axis; uniform and spatially routed profiles lack these temporal atoms. A coverage-preserving $H_4/H_9/H_{20}$ control reaches 0.01317, versus 0.01499 for $H_4/H_4/H_{25}$, at the same parameter count (Section F). These results support temporal frequency coverage under matched parameter budgets without requiring the original capacity ratio. Non-target atoms still differ; the experiments establish neither statistical equivalence nor an optimal allocation among all viable configurations.

A4: localized factors help beyond an exact scale match. The fixed bank excludes all four target scales yet beats H/H throughout, as do the $\pm 25\%$ and $\pm 50\%$ offset arms (Table 19). On the thinnest layer, fixed-bank error is 0.273, versus 0.00137 with exact augmentation; at $\epsilon = 0.008$, the fixed bank instead performs better. Thus, localized factors help beyond exact matching, but accuracy is neither preserved nor monotone in scale mismatch. Unchanged layer shape and orientation make this a scale-sensitivity study, not a transfer test.

Fixed profiles without a manufactured target. To test localized factors beyond exact-target A4, we evaluate a separate source-specified CDR problem without an analytic target. Families, capacities, scale banks, and training configurations are frozen before reference generation and fitting (Section E.1). The bank excludes the new diffusion scale; no configuration is selected using reference errors. Both B-containing profiles reduce global and layer errors relative to H/H at every paired seed (Table 5). Independent three-grid checks preserve global profile ordering on the two finest grids. This supports fixed localized representations on a related non-manufactured CDR problem, without establishing automatic selection or cross-family transfer.

Table 5: Fixed-profile CDR test, separate from the 16-case audit. Relative L_2 error, mean ± sample standard deviation over three paired seeds; 3,000 updates. The layer band is $x > 0.98$. Bold denotes the lowest mean, not statistical significance.

Profile	Parameters	Global error $\downarrow$	Layer error $\downarrow$
B_{64}/H_{32}	6,752	$\mathbf{0.00636 \pm 0.00133}$	0.00752 ± 0.00063
B_{44}/B_{44}	6,720	0.01004 ± 0.00222	$\mathbf{0.00721 \pm 0.00127}$
H_{52}/H_{52}	6,784	0.28627 ± 0.01701	0.22656 ± 0.00477

4.4 COMPLETE-SOLVER PERFORMANCE

Table 6: Complete-solver comparison at 3,000 updates, seeds 42, 2026, 3407. Error is mean ± sample standard deviation; time is mean recorded training time. Profiles retain solver-specific priors and optimization choices; this is neither a matched factor intervention nor an equal-time comparison.

		Baseline		DimSpec-LR		
Case	Comparator	Error ↓	Time (s)	Error ↓	Time (s)	Params
H4	PirateNet	$0.025 \pm 4.78 \times 10^{-3}$	196.3	$\mathbf{1.47 \times 10^{-5} \pm 1.28 \times 10^{-5}}$	49.5	512
A4	MS Fourier	$0.973 \pm 1.81 \times 10^{-3}$	184.2	$\mathbf{1.59 \times 10^{-3} \pm 4.15 \times 10^{-4}}$	224.7	6,752
S2	PirateNet	$8.16 \times 10^{-3} \pm 2.63 \times 10^{-3}$	227.6	$\mathbf{5.75 \times 10^{-3} \pm 7.82 \times 10^{-4}}$	192.7	36,128
K3	MS Fourier	0.116 ± 0.034	219.8	$\mathbf{0.010 \pm 1.23 \times 10^{-3}}$	240.5	2,240

Table 6 compares complete solvers at 3,000 updates. H4, A4, and K3 show substantial accuracy gains, while S2 improves more modestly with overlapping seed variation. All four DimSpec-LR profiles use fewer trainable parameters. These are complete-configuration comparisons, not isolated factor effects, because priors, constraints, and optimization differ. Section F documents the effects of spectral weighting and checkpoint selection on H4 and projected initialization on A3. Training takes longer on A4/K3, so fixed-update gains do not imply uniform compute-efficiency gains. Sections 4.2 and 4.3 provide separate evidence for factor-design attribution.

Table 7 examines the breadth of these gains across all 16 cases: baselines have lower errors on H1/A1 and S1/S3/S4. This seed-42 audit documents applicability and complements the three-seed means; it does not establish a statistical cross-task ranking.

Table 7: Complete 16-case audit, seed 42: DimSpec-LR versus the lowest-error declared baseline; lower relative L_2 errors are bold (field for H/A/K, scalar flux for S). Neural methods use 3,000 updates; RFM uses its physics fit. Solver-specific priors and optimization remain unmatched; these are not population win rates. Three-seed results: Table 6; full matrix: Table 15.

Case	Lowest-error baseline	Baseline error	DimSpec-LR error
Helmholtz			
H1	RFM	$\mathbf{3.72 \times 10^{-3}}$	8.29×10^{-3}
H2	PirateNet	9.80×10^{-4}	$\mathbf{4.16 \times 10^{-4}}$
H3	SPINN	5.03×10^{-3}	$\mathbf{4.49 \times 10^{-4}}$
H4	PirateNet	0.030	$\mathbf{2.62 \times 10^{-5}}$
Anisotropic CDR			
A1	PirateNet	$\mathbf{2.31 \times 10^{-3}}$	0.023
A2	PirateNet	2.22×10^{-4}	$\mathbf{2.44 \times 10^{-6}}$
A3	PirateNet	0.074	$\mathbf{4.23 \times 10^{-3}}$
A4	MS Fourier	0.975	$\mathbf{1.65 \times 10^{-3}}$
Stationary transport			
S1	Vanilla PINN	$\mathbf{3.74 \times 10^{-3}}$	4.46×10^{-3}
S2	Vanilla PINN	7.95×10^{-3}	$\mathbf{6.60 \times 10^{-3}}$
S3	PirateNet	$\mathbf{7.43 \times 10^{-3}}$	0.011
S4	PirateNet	**0.011**	0.048
Klein–Gordon			
K1	PirateNet	4.22×10^{-3}	$\mathbf{1.11 \times 10^{-3}}$
K2	PirateNet	1.75×10^{-3}	$\mathbf{4.75 \times 10^{-4}}$
K3	MS Fourier	0.093	**0.011**
K4	MS Fourier	0.233	**0.022**

What changes with more baseline training? Table 8 trains PirateNet and an additional CP-PINN comparator on H4/A4/K3/S3 at longer budgets with two learning rates each, selecting both the run and checkpoint by physics validation. CP-PINN is distinct from the SPINN configuration in the coverage audit (Section A.3). It is a single-seed sensitivity check, not an equal-compute comparison.

Table 8: Longer-budget sensitivity at seed 42, not an equal-compute comparison. Two learning rates (multipliers 1 and 0.2) per baseline; run and checkpoint selection use physics validation. Cost sums both trials. CP-PINN is a homogeneous neural CP comparator, not a matched dictionary control. S3 uses the main profile at learning rate 0.002 and scalar-flux error.

Case	Baseline	Updates	Extended error	DimSpec-LR (3k)	Both LR trials (s)
A4	CP-PINN	50,000	0.983	1.65×10^{-3}	6380
A4	PirateNet	20,000	0.608	1.65×10^{-3}	4582
H4	CP-PINN	50,000	0.075	2.62×10^{-5}	3607
H4	PirateNet	20,000	1.25×10^{-3}	2.62×10^{-5}	3007
K3	CP-PINN	50,000	7.77×10^{-3}	0.011	8990
K3	PirateNet	20,000	0.632	0.011	5947
S3	CP-PINN	50,000	3.49×10^{-3}	0.011	4077
S3	PirateNet	20,000	0.034	0.011	2618

Budget sensitivity differs across cases. At 50,000 updates, CP-PINN reaches 7.77×10^{-3} on K3 and 3.49×10^{-3} on S3, below DimSpec-LR's 3,000-update errors of approximately 0.011. Short-budget gaps thus do not establish representation barriers. Extended baselines still trail on H4/A4: H4 PirateNet reaches 1.25×10^{-3}; A4 errors remain 0.983 for CP-PINN and 0.608 for PirateNet. Constraints, conditioning, and optimization also differ, preventing a representation-only interpretation of these budget-dependent gaps.

4.5 Applicability and remaining limitations

Curved-field errors improve with higher rank: increasing R from 8 to 32 on H1/A1 reduces H1 error from 0.0136 to 0.00432 and A1 from 0.162 to 0.0139 without changing their factors (Table 20), but neither surpasses its lowest-error baseline. Higher rank also changes capacity and optimization, so the improvement does not identify interaction rank as the sole bottleneck.

S3 shows budget sensitivity, whereas S4's interface dictionary leaves a gap whose cause remains unresolved. The controls do not distinguish interface treatment, angular approximation, rank, and optimization. Known interface locations alone therefore do not ensure competitive accuracy. These are limitations of the tested profiles, not established barriers of CP representations.

5 Conclusion and Limitations

DimSpec-LR combines neural, dictionary-based, and hybrid coordinate factors under a shared CP interaction. Parameter-matched interventions hold this interaction fixed to examine family selection, capacity allocation, and coordinate assignment in finite-budget physics-informed solving.

Three findings emerge. Coordinate correspondence strongly affects accuracy, yet rich homogeneous or capacity-routed alternatives match or outperform prescribed mixed profiles. Covering required directional frequencies improves performance without requiring a unique capacity ratio. Target-mode coverage alone does not determine trained accuracy: representability and finite-budget trainability are distinct. A separate source-specified CDR test also supports fixed localized profiles without an analytic target solution.

The findings concern expert-designed profiles, specified priors, and finite budgets. Solver-specific initialization, preconditioning, and constraints prevent representation-only attribution of complete-solver gains. Curved fields and interfaces remain challenging; transport evaluation concerns scalar flux, not angular-field accuracy. Automatic selection and transfer across unseen PDE families remain outside the scope.

This controlled evaluation approach shows why coordinate suitability, capacity allocation, and architectural heterogeneity should be distinguished when developing and comparing low-rank physics-informed solvers.

AI USE STATEMENT

In this work, we used generative AI tools to assist with several aspects of the research and manuscript preparation. Specifically, AI tools assisted in drafting the Introduction and methodology descriptions, designing ablation study protocols, and generating and editing the associated experimental code. They were also used to create or modify scientific figures and images, improve the manuscript's readability through language editing and polishing, and format bibliographic references. All AI-assisted materials were reviewed by the authors, including checks of the experimental designs, code, scientific figures, and manuscript content for correctness and consistency with the reported results. The authors take full responsibility for the final content of this work, including all scientific claims, experimental results, and AI-assisted artifacts.

REPRODUCIBILITY STATEMENT

Source code for the four PDE benchmark families is provided in the supplementary materials and will be released in the authors' GitHub repository after the anonymous review process. Section 3 defines the shared framework, while Section A specifies factor profiles, hyperparameters, random seeds, and training procedures. Section B describes the PDEs and reference solutions; Sections E and F document the controlled experiments and numerical-reference checks. The assumptions and proof of the approximation bound are given in Section C.

REFERENCES

Jingrun Chen, Xurong Chi, Weinan E, and Zhouwang Yang. Bridging traditional and machine learning-based algorithms for solving pdes: The random feature method, 2022. URL https://arxiv.org/abs/2207.13380.

Junwoo Cho, Seungtae Nam, Hyunmo Yang, Seok-Bae Yun, Youngjoon Hong, and Eunbyung Park. Separable physics-informed neural networks. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 23761–23788. Curran Associates, Inc., 2023. doi: 10.52202/075280-1032. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/4af827e7d0b7bdae6097d44977e87534-Paper-Conference.pdf.

Ingrid Daubechies. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, January 1992. ISBN 9781611970104. doi: 10.1137/1.9781611970104. URL http://dx.doi.org/10.1137/1.9781611970104.

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*, 2021. URL https://arxiv.org/abs/2010.08895.

Dong Liu, Chao Wu, Qi-Long Chen, Kang-Jun He, Jin-Chao Zhang, Qian Zhang, Kai Wang, Chen Zhao, and Yang Liu. Frequency reduction and order elevation (froe): A physics-informed deep learning method for ultra-fine group resonance self-shielding in heterogeneous systems. *Nuclear Engineering and Technology*, 58 (5):104150, May 2026. ISSN 1738-5733. doi: 10.1016/j.net.2026.104150. URL http://dx.doi.org/10.1016/j.net.2026.104150.

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, March 2021. ISSN 2522-5839. doi: 10.1038/s42256-021-00302-5. URL http://dx.doi.org/10.1038/s42256-021-00302-5.

M. Raissi, P. Perdikaris, and G.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, February 2019. ISSN 0021-9991. doi: 10.1016/j.jcp.2018.10.045. URL http://dx.doi.org/10.1016/j.jcp.2018.10.045.

Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 7462–7473. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/53c04118df112c13a8c34b38343b9c10-Paper.pdf.

Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 7537–7547. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/55053683268957697aa39fba6f231c68-Paper.pdf.

Sai Karthikeya Vemuri, Tim Büchner, Julia Niebling, and Joachim Denzler. *Functional Tensor Decompositions for Physics-Informed Neural Networks*, pp. 32–46. Springer Nature Switzerland, December 2024. ISBN 9783031783890. doi: 10.1007/978-3-031-78389-0_3. URL http://dx.doi.org/10.1007/978-3-031-78389-0_3.

Sai Karthikeya Vemuri, Tim Büchner, Julia Niebling, and Joachim Denzler. Scalable and expressive physics-informed neural networks via functional tensor decomposition. *Pattern Recognition Letters*, 207:184–192, September 2026. ISSN 0167-8655. doi: 10.1016/j.patrec.2026.06.027. URL http://dx.doi.org/10.1016/j.patrec.2026.06.027.

Sifan Wang, Bowen Li, Yuhan Chen, and Paris Perdikaris. Piratenets: Physics-informed deep learning with residual adaptive networks. *Journal of Machine Learning Research*, 25(402):1–51, 2024. URL http://jmlr.org/papers/v25/24-0313.html.

Xinling Yu, Sean Hooten, Ziyue Liu, Yequan Zhao, Marco Fiorentino, Thomas Van Vaerenbergh, and Zheng Zhang. Separable operator networks. *Transactions on Machine Learning Research*, 2024. URL https://arxiv.org/abs/2407.11253.

A Reproducibility Details

This appendix specifies the fixed profiles, branch definitions, optimization protocols, baseline architectures, and all benchmark equations used in the paper. Interior reference values are used only for evaluation. Main DimSpec-LR profiles and factor-design interventions use $R = 16$ product components; separate rank checks vary only R. Each formal profile is held fixed across replicated seeds.

A.1 Casewise coordinate profiles

Table 9: Factor profiles for DimSpec-LR. All cases use $R = 16$. S, H, B, T, C, J, P, G and DB4 are defined in Section A.2.

Case	Coordinates	Per-coordinate profile	Trainable part	Intended signature	Params
H1	(x, y)	(H_{32}, H_{32})	coefficients	curved mixed spectrum	4,224
H2	(x, y)	(S_{64}, S_{64})	coefficients	dense off-grid spectrum	2,048
H3	(x, y)	(S_{24}, S_{24})	coefficients	unequal axis frequency	768
H4	(x, y)	(S_{16}, S_{16})	coefficients	finite multimode spectrum	512
A1	(x, y)	(H_{48}, H_{64})	coefficients	curved smooth/local variation	7,296
A2	(x, y)	(S_{48}, S_{48})	coefficients	directional oscillation	1,536
A3	(x, y)	(T^x_{27}, T^y_{27})	coefficients	fixed oblique carrier	1,792
A4	(x, y)	(B_{64}, H_{32})	coefficients	localized layer plus smooth axis	6,752
S1	(x, μ)	$(P(H_{64}; 48, 2), P(C_{32}; 48, 2))$	dictionary + MLP	space–angle mismatch	11,154
S2	(x, μ)	$(G(64, 4), G(64, 4))$	neural	coupled smooth transport	36,128
S3	(x, μ)	$(P(H_{64}; 48, 2), P(C_{32}; 48, 2))$	dictionary + MLP	anisotropic collision modes	11,154
S4	(x, μ)	$(P(J_{64}; 48, 2), P(C_{32}; 48, 2))$	dictionary + MLP	material interfaces	11,250
K1	(a, b, t)	(H_2, H_2, H_4)	coefficients	smooth space, periodic time	704
K2	(a, b, c, t)	(H_2, H_2, H_2, H_4)	coefficients	extra smooth spatial axis	896
K3	(a, b, t)	(H_4, H_4, H_{24})	coefficients	fast time, smooth space	2,240
K4	(a, b, t)	(H_{16}, H_{16}, H_{12})	coefficients	multimode space–time	3,008
N1	(u, s, μ)	(DB4–2048, SIREN$(64, 3)$, analytic)	energy/space	uniform plate	42,256
N2	(u, r, ξ, φ)	(DB4–2048, SIREN$(64, 3)$, analytic, analytic)	energy/space	uniform cylinder	42,256
N3	(u, s, μ)	(DB4–2048, 3-region SIREN$(64, 3)$, analytic)	energy/space	heterogeneous plate	61,232
N4	(u, r, ξ, φ)	(DB4–2048, 3-region SIREN$(64, 3)$, analytic, analytic)	energy/space	heterogeneous cylinder	61,232

The profile notation records factor family, capacity, and coordinate assignment. For example, (B_{64}, H_{32}) in A4 assigns a localized boundary-layer dictionary to the streamwise coordinate and a hybrid smooth/spectral dictionary to the cross-stream coordinate. It does not change the CP contraction. The analytic angular entries in N1–N4 have no trainable parameters. The audit specifies reproducible configurations, not an optimal selection rule. Forcing and prescribed data are allowed inputs, including u_t when specified. The A3 carrier, K3 bandwidth, and K4 spatial modes are benchmark-aware modeling choices, not consequences of the operator alone. Interior reference labels do not enter the loss; held-out profile selection is outside this evaluation.

A.2 Exact branch definitions

For $x \in [\ell, h]$, define the normalized coordinates $z = 2(x - \ell)/(h - \ell) - 1$ and $v = (x - \ell)/(h - \ell)$. Each dictionary is evaluated independently on one coordinate and multiplied by a trainable coefficient matrix with R columns.

Sine, Fourier, and hybrid dictionaries. The sine dictionary is

$$S_m(x) = \big(\sin(j\pi x)\big)_{j=1}^{m}. \tag{7}$$

Table 10: Profile construction audit for all 20 cases. Physical cues explain the design, but do not uniquely determine the recorded capacities. Unless an explicit construction is stated, numbers are fixed expert choices, not outputs of a blind bandwidth-selection rule. Exact profiles are in Table 9.

Case	Available PDE information	Directional cue / family	Recorded capacity or scale construction
H1	Variable index and ring source	Mixed spectrum; H/H	Fixed $32/32$; no reference-based formula is claimed.
H2	Variable reaction and forcing	Off-grid oscillation; S/S	Fixed $64/64$ integer bank; fractional carriers are not inserted.
H3	Forcing with axis frequencies 1,16	Oscillation; S/S	Fixed $24/24$; no per-axis bandwidth search.
H4	Four forcing modes, maximum 16	Multimode; S/S	Fixed $16/16$ covers the declared forcing modes.
A1	Curved velocity and narrow inflow	Mixed scales; H/H	Fixed $48/64$; directional capacity is an expert choice.
A2	Diffusion tensor and forcing modes	Axis oscillation; S/S	Fixed $48/48$; anisotropy alone does not determine frequency.
A3	Rotated tensor and oscillatory forcing	Oblique carrier; T^x/T^y	$\kappa = 12\pi$ is benchmark-specific and visible in forcing, not implied by the tensor; fixed 28-angle frame ($m = 27$).
A4	ε, drift, and forcing	Outflow layer; B/H	Fixed $64/32$; append $s = \varepsilon = 0.0025$ to the common scale bank.
S1	Degree-7 collision kernel	Space / angular moments; $P(H)/P(C)$	Fixed $64/32$, corrections $(48, 2)$; polynomial capacity exceeds kernel degree.
S2	Constant isotropic collision, inflow	Smooth factors; G/G	Fixed width 64, depth 4 on both axes.
S3	$1 + 0.9\mu\mu'$ kernel, inflow	Space / angular moments; $P(H)/P(C)$	Same fixed $64/32$ and corrections as S1, not a degree-1 truncation.
S4	Material edges and isotropic kernel	Interfaces / angle; $P(J)/P(C)$	Edges $0.25, 0.75$ from coefficients; fixed width $a = 0.04$, modes $64/32$.
K1	Forcing and prescribed u, u_t	Polynomial space / periodic time; H	Fixed $2/2/4$; common polynomial block retained.
K2	Forcing and prescribed u, u_t	Extra smooth spatial axis; H	Fixed $2/2/2/4$.
K3	Forcing and time-dependent boundary data	Fast time; H	Fixed $4/4/24$ covers the declared 20-radian carrier; 24 is not uniquely inferred from it.
K4	Initial spatial modes and forcing	Multimode space/time; H	Fixed $16/16/12$; spatial 16π mode is already in prescribed initial data.
N1	Material tables, plate geometry	Multiscale energy / smooth space	DB4 support from coefficient differences and anchors; budget 2,048 and SIREN $(64, 3)$ fixed.
N2	Material tables, cylinder geometry	Same energy/space design	Same support rule and fixed budget as N1; angular factors prescribed.
N3	Piecewise materials and plate interfaces	Material-local spatial factors	Same energy rule; three spatial branches follow the declared regions.
N4	Piecewise materials and radial interfaces	Material-local spatial factors	Same energy rule; three radial branches follow the declared regions.

The periodic Fourier dictionary is

$$F_m(x) = \left[1, \left(\sin(2\pi jv)\right)_{j=1}^{m}, \left(\cos(2\pi jv)\right)_{j=1}^{m}\right]. \tag{8}$$

The hybrid dictionary combines a cubic polynomial block and two frequency conventions,

$$H_m(x) = \left[1, z, z^2, z^3, (\sin(jx), \cos(jx), \sin(j\pi x), \cos(j\pi x))_{j=1}^{m}\right]. \tag{9}$$

Boundary-layer dictionary. Let

$$\mathcal{S}_0 = \{0.00125, 0.0025, 0.005, 0.01, 0.02, 0.04, 0.08, 0.16, 0.32\}. \tag{10}$$

When the operator supplies a diffusion length ε, it is appended to this scale list, retaining duplicates. With $E_s(v) = \exp((v-1)/s)$,

$$B_m(x) = [H_m(x), (E_s(v))_{s\in\mathcal{S}}, (vE_s(v))_{s\in\mathcal{S}}, (v(1 - E_s(v)))_{s\in\mathcal{S}}]. \tag{11}$$

The operator-scale atom is derived from the declared PDE coefficient, not from the reference solution. For A4, $\varepsilon = 0.0025$ is already in $\mathcal{S}_0$. Appending it duplicates the three corresponding atoms, changing the parameterization without enlarging the dictionary span. The scale-sensitivity study in Section E instead uses a modified fixed bank that excludes all tested target scales.

Steerable dictionary. For $j = 0, \dots, m$, set $\theta_j = \pi j/(2m)$, $\kappa = 12\pi$, and $w(x) = (x - \ell)(h - x)/(h - \ell)^2$. The two coordinate frames are

$$T_m^x(x) = [(w(x)\sin(\kappa x \cos\theta_j))_j, (w(x)\cos(\kappa x \cos\theta_j))_j], \tag{12}$$

$$T_m^y(x) = [(w(x)\sin(\kappa x \sin\theta_j))_j, (w(x)\cos(\kappa x \sin\theta_j))_j]. \tag{13}$$

Their CP products synthesize a finite set of oblique carriers while retaining one-dimensional factor evaluation.

Collision and interface dictionaries. The collision dictionary is $C_m(\mu) = (P_j(\mu))_{j=0}^m$, where P_j is the Legendre polynomial of degree j. For material edges $\mathcal{E} = \{0.25, 0.75\}$ and width $a = 0.04$, the spatial interface dictionary is

$$J_m(x) = \left[H_m(x), \left(\max(v - e, 0), e^{-|v-e|/a}, \tanh((v - e)/a)\right)_{e\in\mathcal{E}}\right]. \tag{14}$$

Dictionary with neural correction. A corrected factor $P(B; w, d)$ consists of $B(x)C$ plus $\mathrm{softplus}(\gamma)n(x)$, where n is a d-hidden-layer tanh network of width w. The scalar γ is trainable and initialized to -2. The correction output has dimension R.

Gated neural factor. For $G(w, d)$, two embeddings U and V and a hidden state h_0 are produced from the normalized coordinate. Each layer computes

$$g_j = \tanh(W_j h_{j-1} + b_j), \qquad h_j = (1 - g_j) \odot U + g_j \odot V, \tag{15}$$

followed by an R-dimensional linear output. All affine parameters are trainable.

Sparse multiscale DB4 factor. The energy branch starts with 128 coarse DB4 coefficients and seven detail levels. The numbers of active detail coefficients from coarse to fine are $(128, 256, 512, 530, 298, 129, 67)$, giving 2,048 trainable coefficients per rank component. Standard DB4 synthesis produces a 16,384-point signal, followed by continuous cubic interpolation in lethargy. The value at the high-energy entrance is subtracted, so the learned correction is zero there. Seventy-five percent of active detail locations are uniform anchors. The remainder are selected from a fixed indicator built from first and second differences of the material cross sections. No reference flux enters this support selection.

A.3 Optimization and baseline definitions

Table 11: Main DimSpec-LR training configurations. Batch entries denote collocation samples unless a quadrature grid is shown. For N1–N4, counts refer to space–direction samples, including opposite-direction pairs (not energy rows). Uniform-material N1–N2 use 8 training pairs and 16 validation pairs; three-region N3–N4 use 9 pairs for each. Each sample spans the active energy prefix in training and all 17,600 rows in validation. Matched interventions have separate protocols; references never enter training or checkpoint selection.

Cases	Updates	Training batch	Physics validation	Initial LR	Check interval	Constraint treatment
H1, A1	3,000	2,048	4,096	2×10^{-3}	100	hard Dirichlet lift
H2	3,000	2,048	2,048	3×10^{-3}	100	hard homogeneous boundary
H3–H4	3,000	2,048	2,048	2×10^{-2}	100	hard homogeneous boundary
A2	3,000	2,048	4,096	10^{-3}	100	sine-enforced zero boundary
A4	3,000	2,048	4,096	2×10^{-3}	100	soft boundary penalty, weight 10
A3	3,000	2,048	4,096	10^{-3}	100	PDE-projected initialization; hard boundary
S1–S4	3,000	64×32 quadrature	128×32	2×10^{-3}	100	half-range inflow; interface sampling
K1	3,000	2,048	2,048	2×10^{-2}	100	initial/boundary constraints
K2	3,000	2,048	2,048	2×10^{-2}	100	initial/boundary constraints
K3	3,000	2,048	2,048	10^{-2}	100	initial/boundary constraints
K4	3,000	2,048	2,048	5×10^{-3}	100	initial/boundary constraints
N1–N2	10,000	16 space–direction	32 space–direction	2×10^{-3} energy	100	restricted scalar/energy entrance form
N3–N4	10,000	18 space–direction	18 space–direction	2×10^{-3} energy	100	restricted scalar/energy entrance form

Table 12: Baseline architectures and their role in the comparison. Neural baselines solve the same casewise PDEs under the same nominal update budget as DimSpec-LR. Ansatz-dependent constraint enforcement, preconditioning, and initialization are not fully matched in complete-solver comparisons; see Section A.3.

Method	Implemented architecture	Comparison role
Vanilla PINN	Joint-coordinate tanh MLP, width 128 and four hidden layers	Standard nonseparable physics-informed network
MS Fourier	Joint Fourier encoding followed by the same width-128 tanh backbone	Generic multiscale spectral baseline
SPINN	One coordinate MLP per axis, rank 32, width 64 and three hidden layers	Homogeneous separable neural baseline
PirateNet	Width-256 adaptive residual backbone with three trainable blocks and Fourier embedding	High-capacity modern PINN baseline
RFM	2,048 localized partition-of-unity random features with a linear physics fit	Frozen-feature, low-optimization-cost baseline

The longer-budget CP-PINN comparator uses rank 32 and one tanh MLP per coordinate, with three hidden layers of width 32 and Xavier-normal weights. Its CP sum is unnormalized, unlike the $R^{-1/2}$ normalization and width-64 branches of the SPINN comparator. It is used in Table 8, not among the five comparators in the complete coverage matrix.

Adam is used for all neural and coefficient-based gradient runs. Gradients are clipped at 10 for A/S/N and at 100 for H/K. H1 and A1 use cosine or two-stage decay to a minimum factor of 0.05. H2/A2/A3/K2 use cosine decay to 0.01 of the initial rate; their rates are in Table 11. H3/H4/K1/K3/K4 use constant rates, and transport profiles use cosine decay. Neutron models use separate initial rates of 2×10^{-3} for energy coefficients and 2×10^{-4} for spatial networks, with a plateau reduction after the full energy grid is active. Computation uses PyTorch 2.5.0 on NVIDIA GeForce RTX 4090 GPUs and Intel Xeon Platinum 8358 CPUs (Section A.4); reported times are not cross-device speed comparisons.

A3 uses a physics-only projected initialization. The declared operator is applied to all products of the fixed one-dimensional atoms on a 96×96 grid. A ridge-regularized coefficient solve followed by a rank-16 truncated SVD initializes the CP factors. This step uses the forcing and differential operator, but not the reference field. The solve uses float32 arithmetic and ridge coefficient 10^{-4}. Because the projection is ill-conditioned, its numerical output is sensitive to device and reduction arithmetic. The baseline configurations do not include an equivalent projected initialization, so A3 evaluates this complete solver configuration.

H/A/S/K methods solve the same casewise PDE; hard and soft constraint implementations depend on the ansatz. In A4 the dictionary arms share a soft boundary penalty, while baseline ansatzes in complete-solver comparisons may enforce it exactly. Neutron models use the restricted objective specified below. Because RFM fits frozen features through its prescribed linear physics solve, it has no meaningful optimizer-update count; its measured fitting time is still reported in the resource audit.

A.4 Hyperparameters, replication, and resources

The reported H2/A2/A3/K2 configurations use rank 16, 3,000 Adam updates, and cosine decay to 0.01 of the initial learning rate. Initial rates are 0.003, 0.001, 0.001, and 0.02, respectively. Physics validation is evaluated every 100 updates; the checkpoint minimizes the validation objective. Candidate rates are $\{0.001, 0.003, 0.01\}$ for A2, $\{0.003, 0.01, 0.02\}$ for H2/K2, and $\{10^{-5}, 10^{-4}, 10^{-3}\}$ for A3. Selection within each candidate set uses physics validation, not reference error. Factors and capacities are fixed throughout. The casewise configurations do not constitute a uniform hyperparameter-selection benchmark.

Table 13 reports three seeds with fixed configurations. The comparator column contains the seed-42 baseline, so this table measures configuration stability rather than a paired multi-seed solver ranking. Reference evaluations are separate from training and checkpoint selection.

For A2/K2, three-candidate configuration-selection costs are 53.5/947.8 seconds on shared RTX 4090 GPUs; for H2/A3, they are 85.3/221.1 seconds on four-thread Xeon Platinum 8358 CPUs. These costs are excluded from the per-fit times. H2/A3 and the seed-2026/3407 replicas use CPU execution, with two threads for replicas. H2/A2/A3/K2 per-fit intervals include initialization, training, and physics validation, but exclude reference evaluation. Other resource entries follow their recorded training intervals. Hardware load and interval definitions differ, so the resource table is descriptive rather than an equal-time speed comparison.

Table 13: Configuration stability across seeds with fixed factors and learning rates. The baseline column gives the seed-42 comparator, not a replicated baseline.

Case	Seed 42	Seed 2026	Seed 3407	Baseline (42)
A2	2.44×10^{-6}	2.19×10^{-6}	2.22×10^{-6}	2.22×10^{-4}
K2	4.75×10^{-4}	5.20×10^{-4}	4.36×10^{-4}	1.75×10^{-3}
H2	4.16×10^{-4}	5.12×10^{-4}	4.36×10^{-4}	9.80×10^{-4}
A3	4.23×10^{-3}	3.91×10^{-3}	3.90×10^{-3}	0.074

B Benchmark Definitions

For manufactured cases, the forcing is $f = \mathcal{N}[u^\star]$ and prescribed boundary or initial data are taken from $u^\star$. The field is not used as an interior training target. H1 and A1 are instead specified by

coefficients, source, and boundary data; their reference fields are produced by independent finite-difference solves.

B.1 H1–H4: HELMHOLTZ

H1 is the numerical-reference curved-wave case on $\Omega = (-1, 1)^2$:

$$\Delta u + k_0^2 n(x, y)^2 u = s(x, y), \qquad u|_{\partial\Omega} = 0, \qquad k_0 = 5, \tag{16}$$

where

$$n(x, y) = 1 + 0.35 \exp\left[-\frac{(x - 0.25)^2 + (y + 0.15)^2}{0.12}\right] + 0.15 \sin(2\pi xy), \tag{17}$$

$$s(x, y) = \exp\left[-120\left(\sqrt{(x + 0.45)^2 + (y - 0.15)^2} - 0.18\right)^2\right]. \tag{18}$$

The reported 513×513 finite-difference reference differs from the nested 257×257 solution by 0.066% in relative L_2.

H2 is the off-dictionary case on $(0, 1)^2$:

$$\Delta u + q(x, y)u = f, \qquad u|_{\partial\Omega} = 0, \tag{19}$$

$$q(x, y) = 1 + 0.75 \sin(2\pi x) \cos(2\pi y), \tag{20}$$

$$u^\star = x(1 - x)y(1 - y) \left[\sin(9.5\pi x) \sin(3.75\pi y) + 0.35 \sin(4.25\pi x) \sin(12.5\pi y)\right]. \tag{21}$$

H3 and H4 solve $\Delta u + u = f$ on $[0, 1]^2$ with zero Dirichlet data:

$$\text{H3}: \quad u^\star = \sin(\pi x) \sin(16\pi y), \tag{22}$$

$$\text{H4}: \quad u^\star = \sum_{j=1}^{4} \frac{1}{j} \sin(a_j \pi x) \sin(b_j \pi y), \tag{23}$$

$$\{(a_j, b_j)\}_{j=1}^4 = \{(1, 16), (2, 8), (4, 4), (8, 2)\}. \tag{24}$$

B.2 A1–A4: ANISOTROPIC CONVECTION–DIFFUSION–REACTION

All cases use $\Omega = (0, 1)^2$ and

$$-\nabla \cdot (D\nabla u) + \boldsymbol{\beta} \cdot \nabla u + cu = f. \tag{25}$$

A1 is the numerical-reference curved-streamline case:

$$D = 10^{-2} I, \qquad c = 0.1, \qquad f = 0, \tag{26}$$

$$\boldsymbol{\beta}(x, y) = \left(\sin(\pi x) \cos(\pi y), -\cos(\pi x) \sin(\pi y)\right). \tag{27}$$

The left boundary is

$$u(0, y) = e^{-150(y-0.25)^2} - (1 - y)e^{-150(0.25)^2} - ye^{-150(0.75)^2}, \tag{28}$$

and the other three traces are zero. The 513×513 reference differs from the nested 257×257 solution by 0.069%.

A2–A4 use homogeneous Dirichlet data and manufactured forcing:

$$\text{A2}: \quad D = \operatorname{diag}(16^{-2}, 1), \quad \boldsymbol{\beta} = 0, \quad c = 0, \quad u^\star = \sin(16\pi x) \sin(\pi y), \tag{29}$$

$$\text{A3}: \quad D = R_{\pi/4} \operatorname{diag}(1, 0.01) R_{\pi/4}^\top, \quad \boldsymbol{\beta} = (0.25, -0.10), \quad c = 0.5, \tag{30}$$

$$u^\star = 32x(1 - x)y(1 - y) \sin\big(12\pi(x \cos\tfrac{\pi}{4} + y \sin\tfrac{\pi}{4})\big), \tag{31}$$

$$\text{A4}: \quad D = \varepsilon I, \quad \varepsilon = 2.5 \times 10^{-3}, \quad \boldsymbol{\beta} = (1, 0.25), \quad c = 0, \tag{32}$$

$$u^\star = x \frac{1 - e^{(x-1)/\varepsilon}}{1 - e^{-1/\varepsilon}} \sin(\pi y). \tag{33}$$

B.3 S1–S4: STATIONARY SLAB TRANSPORT

For $(x, \mu) \in (0, 1) \times (-1, 1)$,

$$\mu \partial_x \psi + \Sigma_t(x)\psi = \frac{\Sigma_s(x)}{2} \int_{-1}^{1} p(\mu, \mu')\psi(x, \mu')\, d\mu', \qquad \Sigma_t = \Sigma_s + \Sigma_a. \tag{34}$$

The inflow is $\psi(0, \mu) = 1$ for $\mu > 0$ and $\psi(1, \mu) = 0$ for $\mu < 0$. The four coefficient sets are

$$\text{S1}: \quad \Sigma_s = x, \quad \Sigma_a = 0, \quad p = \sum_{\ell=0}^{7} d_\ell P_\ell(\mu) P_\ell(\mu'), \tag{35}$$

$$(d_0, \ldots, d_7) = (1, 1.98398, 1.50823, 0.70075, 0.23489, 0.05133, 0.00760, 0.00048), \tag{36}$$

$$\text{S2}: \quad \Sigma_s = 0.8, \quad \Sigma_a = 0.2, \quad p = 1, \tag{37}$$

$$\text{S3}: \quad \Sigma_s = 3, \quad \Sigma_a = 0.3, \quad p = 1 + 0.9\mu\mu', \tag{38}$$

$$\text{S4}: \quad (\Sigma_s, \Sigma_a) = \begin{cases} (1, 0.5), & 0 < x < 0.25, \\ (8, 0.1), & 0.25 \le x < 0.75, \\ (1, 0.5), & 0.75 \le x < 1, \end{cases} \qquad p = 1. \tag{39}$$

The reported observable is the scalar flux $\phi(x) = \int_{-1}^{1} \psi(x, \mu)\, d\mu$. Evaluation uses an independent discrete-ordinates reference with 2,048 spatial cells and 128 angular nodes, distinct from the training quadrature.

B.4 K1–K4: KLEIN–GORDON

The common equation is

$$u_{tt} - \sum_{j=1}^{d} u_{x_j x_j} + u^2 = f, \qquad f = \mathcal{N}[u^\star], \tag{40}$$

with initial and spatial boundary data from $u^\star$. The cases are

$$\text{K1}: \quad u^\star(a, b, t) = (a + b)\cos(2t) + ab\sin(2t), \qquad (a, b, t) \in [-1, 1]^2 \times [0, 10], \tag{41}$$

$$\text{K2}: \quad u^\star(a, b, c, t) = (a + b + c)\cos t + abc \sin t, \qquad (a, b, c, t) \in [-1, 1]^3 \times [0, 10], \tag{42}$$

$$\text{K3}: \quad u^\star(a, b, t) = (a + b)\cos(20t) + ab\sin(20t), \qquad (a, b, t) \in [-1, 1]^2 \times [0, 2], \tag{43}$$

$$\text{K4}: \quad u^\star(a, b, t) = \sin(16\pi a)\sin(\pi b)\cos(2t) \tag{44}$$

$$+ \tfrac{1}{4}\sin(4\pi a)\sin(8\pi b)\sin(6t), \quad (a, b, t) \in [0, 1]^2 \times [0, 2]. \tag{45}$$

B.5 N1–N4: RESTRICTED NEUTRON-MODERATION APPLICATION

Let $u = \log(E_{\text{ref}}/E)$ denote lethargy and x the spatial coordinate. The FROE formulation represents an angular primitive I (Liu et al., 2026): $\psi_{\text{p}} = \partial_\mu I_{\text{p}}$ in a plate and $\psi_{\text{c}} = \partial_\xi \partial_\varphi I_{\text{c}}$ in a cylinder. The normalized scalar observables are

$$\phi_{\text{p}}(u, s) = \tfrac{1}{2} I_{\text{p}}(u, s, 1), \qquad \phi_{\text{c}}(u, r) = \tfrac{1}{4\pi} I_{\text{c}}(u, r, 1, 2\pi). \tag{46}$$

Isotropic scattering makes the source, not necessarily ψ, direction-independent.

For isotope k with mass number A_k, define

$$\alpha_k = \left(\frac{A_k - 1}{A_k + 1}\right)^2, \qquad \Delta u_k = \log(1/\alpha_k). \tag{47}$$

For $\alpha_k > 0$, the slowing-down source per unit angular measure is

$$S_k[\phi](u, x) = e^{-u} \int_{u - \Delta u_k}^{u} \frac{N_k(x)\sigma_{s,k}(u', x)}{1 - \alpha_k} \phi(u', x) e^{u'}\, du'. \tag{48}$$

Hydrogen has $\alpha_H = 0$ and a semi-infinite energy history. Splitting at the highest-energy boundary u_0 gives

$$S_H[\phi](u,x) = e^{-u}\left[C_H^{\mathrm{ext}}(x) + \int_{u_0}^{u} N_H(x)\sigma_{s,H}(u',x)\phi(u',x)e^{u'}\,du'\right], \tag{49}$$

where C_H^{ext} is fixed by the prescribed high-energy entrance closure.

The underlying angular transport equation is

$$\mathcal{L}_g\psi_g + \Sigma_t^{m(x)}(u)\psi_g - \sum_{k\in\mathcal{K}_{m(x)}} S_k[\phi_g] = 0, \tag{50}$$

with streaming operators

$$\mathcal{L}_{\mathrm{p}}\psi = \mu\,\partial_s\psi, \tag{51}$$

$$\mathcal{L}_{\mathrm{c}}\psi = \sqrt{1-\xi^2}\left(\cos\varphi\,\partial_r\psi - \frac{\sin\varphi}{r}\partial_\varphi\psi\right). \tag{52}$$

Full transport requires angular-flux continuity across material interfaces, vacuum inflow at exterior boundaries, and cylindrical regularity at $r = 0$. The reported runs use the restricted ansatz and reduced objective below.

The implementation trains a scalar CP core

$$\widehat{\phi}_g(u,x) = \phi_{0,g}(u,x) + R^{-1/2}\sum_{q=1}^{R} E_q(u)X_q(x), \tag{53}$$

with fixed analytic angular factors $\widehat{I}_{\mathrm{p}} = (\mu+1)\widehat{\phi}_{\mathrm{p}}$ and $\widehat{I}_{\mathrm{c}} = (\xi+1)\varphi\widehat{\phi}_{\mathrm{c}}$. Thus $\widehat{\psi}_g = \widehat{\phi}_g$ is direction-independent; this is a trial-space restriction, not a consequence of isotropic scattering. The spatial ansatz imposes zero scalar flux at exterior boundaries, stronger than vacuum inflow; the cylinder has no separate hard $\partial_r\phi(u,0) = 0$ constraint. Energy and spatial factors use sparse DB4 and SIREN, respectively. The four cases use the same 17,600 lethargy points: N1 is a uniform UO plate, N2 a uniform UO cylinder, N3 a three-region HO–UO–HO plate, and N4 a three-region O–UO–O cylinder. The plate interfaces are at one-quarter and three-quarters of the slab; the cylinder interfaces are $r = 1$ and $r = 2$ in a radius-three domain. Cross sections and number densities are read from the corresponding ultra-fine-group material tables.

Implemented reduced objective. Write $T_g = \kappa_g\partial_x\widehat{\phi}_g$, $C_g = \Sigma_t\widehat{\phi}_g$, $S_g = \sum_k S_k[\widehat{\phi}_g]$, with $\kappa_{\mathrm{p}} = \mu$ and $\kappa_{\mathrm{c}} = \sqrt{1-\xi^2}\cos\varphi$. For empirical means on sample set B, let $d_B(Z_1,\dots,Z_j) = \max(10^{-10}, \langle\sum_i Z_i^2\rangle_B)$. For the full batch, write $d(Z) = \max(10^{-10}, \langle Z^2\rangle)$. Training and checkpoint selection use

$$\mathcal{J}_N = \frac{1}{|\mathcal{B}|}\sum_{B\in\mathcal{B}}\left[\frac{\langle(T_g + C_g - S_g)^2\rangle_B}{d_B(T_g, C_g, S_g)} + \frac{\langle(C_g - S_g)^2\rangle_B}{d_B(C_g, S_g)}\right] + 0.1\frac{\langle\max(-\widehat{\phi}_g, 0)^2\rangle}{d(\widehat{\phi}_g)}. \tag{54}$$

Here $\mathcal{B}$ separates material groups when present. Denominators are recomputed from the current field with stopped gradients, rather than fixed operator scales. Opposite angular samples cancel the streaming cross term, leaving a source-mismatch and spatial-gradient penalty. This restricted least-squares objective does not certify the full angular transport solution; scalar-reference agreement is the reported application metric.

Table 14: N2 scalar-flux application at 10,000 outer updates, three seeds (scope in Section B.5). Error is mean ± sample standard deviation; time is mean end-to-end seconds. Both models use the restricted application pipeline, not an unrestricted angular-transport residual.

			Comparator		DimSpec-LR		
Application	Case	Model	Scalar error	Time (s)	Scalar error	Time (s)	Params
Restricted neutron	N2	MS Fourier	0.806 ± 0.011	2247.8	$\mathbf{0.019\pm7.57\times10^{-4}}$	1234.0	42,256

N1–N4 remain separate from the 16-case PDE solver suite.

C Approximation Perspective

This section proves Equation (5) and clarifies what the fixed component count represents. For a single product,

$$\prod_{d=1}^{D} f_d - \prod_{d=1}^{D} b_d = \sum_{d=1}^{D} (f_d - b_d) \prod_{j<d} b_j \prod_{j>d} f_j. \tag{55}$$

On a product domain, the L^2 norm of each separated term factors into the product of one-dimensional norms. Applying the triangle inequality, the factor error assumptions, and the L^∞ bound on q yields Equation (5) after summing over the R components.

The quantity δ_R is the interaction error of the chosen CP component count. The terms $\varepsilon_{d,r}$ are the errors of the coordinate factors. Changing a branch can reduce the latter without reducing the former. Curved features may still require larger R. This is a possible contributor to the H1/A1 errors, not a decomposition of their measured training errors.

The rank in Equation (2) counts only the trainable CP core. If the hard multiplier q has a separated representation with Q terms and the lift has L terms, expanding their products gives the upper bound

$$\mathrm{rank}_{\mathrm{CP}}(\widehat{u}) \le L + QR. \tag{56}$$

Thus, a matrix-rank diagnostic of a reference field is informative but need not equal the effective rank of the lifted model. The approximation statement also does not imply a bound on derivatives or on the strong-form PDE residual; those require operator-specific regularity and stability assumptions.

D Complete Results

Table 15: Complete finite-budget single-seed audit: 16 PDE cases and four restricted scalar-flux applications. Casewise configurations are specified in Table 11 and Section A.4. Neural methods use the common casewise update count; RFM uses its prescribed physics fit. N1–N4 measure scalar-flux agreement within the scope of Section B.5, not full angular transport.

Family	Case	Vanilla PINN	MS Fourier	SPINN	PirateNet	RFM	DimSpec-LR
Helmholtz	H1	0.503	0.852	0.347	0.043	$\mathbf{3.72\times10^{-3}}$	8.29×10^{-3}
	H2	0.013	1.099	1.109	9.80×10^{-4}	0.098	$\mathbf{4.16\times10^{-4}}$
	H3	3.817	0.038	5.03×10^{-3}	0.016	0.409	$\mathbf{4.49\times10^{-4}}$
	H4	2.077	0.143	0.468	0.030	0.344	$\mathbf{2.62\times10^{-5}}$
Anisotropic CDR	A1	0.065	1.712	0.114	$\mathbf{2.31\times10^{-3}}$	0.099	0.023
	A2	1.000	1.000	1.000	2.22×10^{-4}	0.731	$\mathbf{2.44\times10^{-6}}$
	A3	2.395	0.353	1.001	0.074	0.219	$\mathbf{4.23\times10^{-3}}$
	A4	0.984	0.975	0.984	0.978	1.007	$\mathbf{1.65\times10^{-3}}$
Stationary transport	S1	$\mathbf{3.74\times10^{-3}}$	0.092	5.84×10^{-3}	0.014	0.022	4.46×10^{-3}
	S2	7.95×10^{-3}	0.136	0.012	0.010	0.024	$\mathbf{6.60\times10^{-3}}$
	S3	0.013	0.225	0.011	$\mathbf{7.43\times10^{-3}}$	0.077	0.011
	S4	0.094	0.233	0.039	**0.011**	0.218	0.048
Klein–Gordon	K1	0.075	0.048	0.013	4.22×10^{-3}	0.750	$\mathbf{1.11\times10^{-3}}$
	K2	0.013	9.38×10^{-3}	6.87×10^{-3}	1.75×10^{-3}	0.423	$\mathbf{4.75\times10^{-4}}$
	K3	0.704	0.093	0.605	0.947	1.535	**0.011**
	K4	1.033	0.233	1.192	0.650	1.250	**0.022**
Restricted scalar-flux applications; not full angular-transport solutions							
Restricted neutron application	N1	0.254	0.601	0.603	0.468	0.321	**0.048**
	N2	0.666	0.801	0.855	0.745	1.000	**0.020**
	N3	0.720	0.311	0.918	0.685	0.809	**0.080**
	N4	0.752	0.451	0.908	0.603	1.000	**0.107**

Table 16: Single-seed resource audit against the lowest-error declared baseline. Times cover the recorded training intervals. H2/A2/A3/K2 include initialization and physics validation, excluding reference evaluation and configuration selection; † denotes CPU execution. GPU runs use shared hardware. See Section A.4 for resource accounting; these mixed-condition times do not establish speedups. N-family evaluation scope is specified in Section B.5.

			Lowest-error declared baseline			DimSpec-LR		
Family	Case	Method	Error	Time	Params	Error	Time	Params
Helmholtz	H1	RFM	3.72×10^{-3}	3.3	2,049	8.29×10^{-3}	224.1	4,224
	H2	PirateNet	9.80×10^{-4}	215.4	727,045	4.16×10^{-4}	$28.6^{\dagger}$	2,048
	H3	SPINN	5.03×10^{-3}	100.3	21,056	4.49×10^{-4}	52.6	768
	H4	PirateNet	0.030	185.6	727,045	2.62×10^{-5}	51.4	512
Anisotropic CDR	A1	PirateNet	2.31×10^{-3}	280.4	727,045	0.023	171.6	7,296
	A2	PirateNet	2.22×10^{-4}	211.8	727,045	2.44×10^{-6}	19.8	1,536
	A3	PirateNet	0.074	351.7	727,045	4.23×10^{-3}	$72.2^{\dagger}$	1,792
	A4	MS Fourier	0.975	181.1	82,817	1.65×10^{-3}	225.5	6,752
Stationary transport	S1	Vanilla PINN	3.74×10^{-3}	80.4	50,049	4.46×10^{-3}	177.6	11,154
	S2	Vanilla PINN	7.95×10^{-3}	104.4	50,049	6.60×10^{-3}	195.5	36,128
	S3	PirateNet	7.43×10^{-3}	222.4	727,045	0.011	180.7	11,154
	S4	PirateNet	0.011	246.2	727,045	0.048	188.7	11,250
Klein–Gordon	K1	PirateNet	4.22×10^{-3}	466.1	727,173	1.11×10^{-3}	236.1	704
	K2	PirateNet	1.75×10^{-3}	561.0	727,301	4.75×10^{-4}	338.0	896
	K3	MS Fourier	0.093	224.6	68,609	0.011	232.9	2,240
	K4	MS Fourier	0.233	225.1	62,465	0.022	283.5	3,008
Restricted neutron application	N1	Vanilla PINN	0.254	1305.1	25,603	0.048	1199.9	42,256
	N2	Vanilla PINN	0.666	1790.2	25,667	0.020	1243.8	42,256
	N3	MS Fourier	0.311	3052.7	63,494	0.080	1958.6	61,232
	N4	MS Fourier	0.451	3829.8	67,653	0.107	1888.7	61,232

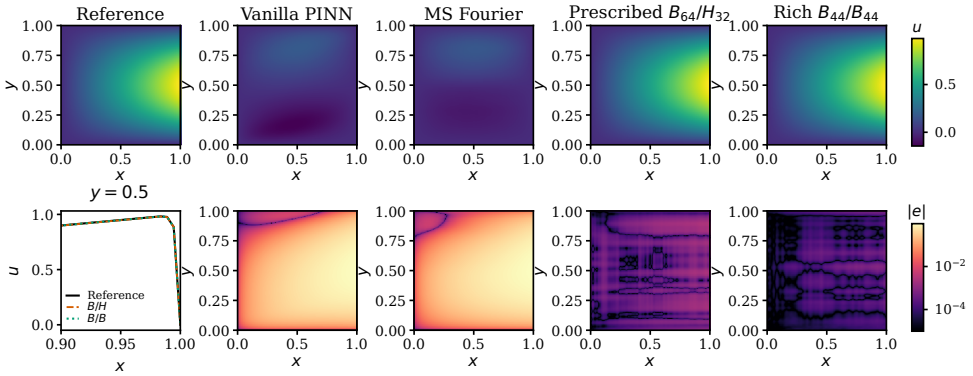


Figure 2: A4 at seed 42, using physics-selected checkpoints. Top: reference and predictions from Vanilla PINN, MS Fourier, prescribed B_{64}/H_{32}, and rich homogeneous B_{44}/B_{44}, with a shared field scale. Bottom: outflow-layer cut at $y = 0.5$ and absolute-error maps on a shared logarithmic scale. Both B-containing profiles recover the layer; accurate coordinate-aligned representations do not require heterogeneous families.

D.1 Details of the main-text controls

The S3 study uses a common learning rate 5×10^{-4}, with 64 spatial samples, 32 angular nodes, and the same initialization stream, inflow treatment, and 3,000-update budget across arms. It is not pooled with the main S3 profile at learning rate 0.002. The A4 rich control gives both axes the H block and localized atoms in B; at matched total size it uses B_{44}/B_{44}, not the untruncated B_{64} union on each axis. A4 keeps dictionary initialization seed 20260728 fixed; its three training seeds vary the stochastic sampling stream. The rich A4 arm was run on a shared GPU; its time is excluded from

comparative cost claims. Table 3 reports its accuracy and size without pooling it with the prescribed profile in the main comparison.

Table 17: Paired control/prescribed error ratios at the same training seed, using the records in Table 3. Values below one favor the control. This secondary analysis includes all nine non-prescribed controls over three seeds, not independent PDE instances or a population significance test.

Case	Control profile	Seed 42	Seed 2026	Seed 3407
A4	H_{52}/H_{52}	520.476	433.928	745.409
A4	H_{72}/H_{31}	519.268	433.272	732.027
A4	H_{32}/B_{64}	519.942	431.419	743.421
A4	B_{44}/B_{44}	0.236	0.299	0.206
S3	H_{36}/H_{36}	3.370	2.775	3.167
S3	H_{64}/H_{7}	0.975	0.915	0.760
S3	C_{32}/H_{64}	6.202	6.665	6.772
S3	C_{32}/C_{32} (62/62)	1.073	1.034	1.027
S3	C_{32}/C_{32} (74/48)	1.006	1.033	0.943

Ratios pair control and prescribed errors by training seed without reselecting checkpoints, showing the variation behind the aggregate means.

The extended runs in Table 8 retain each baseline's casewise loss and constraints, with learning-rate multipliers 1 and 0.2. Run and checkpoint selection use physics validation; reported search costs include both trials.

E Intervention Protocols and Sensitivities

Scope and shared training. The fixed intervention protocol uses four arms and three seeds (42, 2026, 3407) for H4 and K3, and eight arms at four diffusion scales with seed 42 for A4. These H4/K3/A4 studies target existing benchmark families, not held-out profile-selection tests; a separate CDR configuration-transfer test is specified below. Each fit uses 3,000 Adam updates, rank 16, 2,048 fresh interior points per update, 4,096 fixed physics-validation points, and 16,384 evaluation points. Validation is performed every 100 updates, including initialization. The checkpoint minimizes physics validation, never reference error. Within a case, all arms share sampling seeds, constraints, loss weights, learning rate, and gradient clipping.

The loss is the squared strong residual divided by the fixed mean squared forcing on the validation set, plus ten times any soft constraint loss. The forcing normalizer is detached. Constant learning rates are 0.002 on H4/A4 and 0.01 on K3; clipping norms are 100 on H4/K3 and 10 on A4. K3 uses 1,024 points in its shared boundary/initial constraint routine; A4 uses 256 points on each of four edges. Sampling streams are determined by the shared seeds; every fit runs for the full update budget without early stopping.

H4: changing family with a common boundary projection. Every branch is projected by $\Pi b(x) = b(x) - (1-x)b(0) - xb(1)$ on $[0,1]$, imposing the same homogeneous endpoint conditions without multiplying an extra envelope onto sine factors. The dictionaries are S_{70}, H_{17}, and $SC_{35} = [(\sin(j\pi x), \cos(j\pi x))_{j=1}^{35}]$. The first two affine atoms of H_{17} are omitted because Π annihilates them. Each projected dictionary therefore has 70 nonzero atoms per axis and 2,240 trainable coefficients in total. Every dictionary contains the H4 forcing's highest spatial mode. The neural control uses a tanh network with widths 1–62–16 on each axis, omitting the output bias that the projection would annihilate: 2,232 parameters, a 0.36% difference. Dictionary coefficients have Gaussian scale $0.005/\sqrt{m_d}$; neural output weights are scaled by 0.05 after standard initialization. This experiment uses an unpreconditioned strong residual and a larger common parameter budget than the main S_{16}/S_{16} solver, whose loss includes spectral preconditioning. Its absolute errors are not pooled with those of the main solver.

H4: column-amplitude control. For S_{70} and projected H_{17}, a separate paired study compares the raw columns with unit-RMS columns. With $\varphi_j = \Pi\Phi_j$, define

$$s_j = \left(\frac{1}{4096} \sum_{i=1}^{4096} \varphi_j((i - \tfrac{1}{2})/4096)^2 \right)^{1/2}, \qquad \widetilde{\varphi}_j = \varphi_j / s_j, \qquad \widetilde{c}_{j,r}^{(0)} = s_j c_{j,r}^{(0)}.$$

Positive scales preserve the span; rescaled coefficients preserve initial fields and residuals up to roundoff. Basis-only scales are computed in float64 before float32 training. All four cells use the shared H4 protocol, 2,240 parameters, and seeds 42, 2026, and 3407 on the same two-thread Xeon Platinum 8358 CPU (12 fits). This tests diagonal amplitude scaling, not Gram whitening or derivative normalization. Table 18 reports the complete block separately from the GPU family comparison.

Table 18: H4 column-RMS control: relative L_2 mean $\pm$ sample standard deviation over seeds 42, 2026, and 3407. All cells use the same CPU protocol; each has 2,240 parameters and 3,000 updates. Initial fields are paired within each family.

Factors	Raw columns	Unit-RMS columns
S_{70}/S_{70}	0.4804 ± 0.1043	0.4324 ± 0.0932
H_{17}/H_{17}	0.0273 ± 0.0119	0.0118 ± 0.0013

K3: changing capacity with the family held fixed. The profiles (H_4, H_4, H_{25}), (H_{11}, H_{11}, H_{11}), (H_{25}, H_4, H_4), and (H_4, H_{25}, H_4) all contain 144 atoms and 2,304 coefficients. The largest factor uses mode 25, rather than the main profile's mode 24, to match three equal complete H dictionaries exactly. The uniform arm is already the rich homogeneous alternative at this budget; it is not a different-family baseline. Constraints remain soft and shared. Initialization uses Gaussian scale $0.5/\sqrt{m_d}$ with a seed keyed to family, capacity, and repeated occurrence; permutation moves the same initialized coefficient arrays to their new coordinates.

A4: removing exact scale information. At $\epsilon \in \{0.0015, 0.0025, 0.004, 0.008\}$, the fixed scale list is

$$\widetilde{\mathcal{S}}_0 = \{0.00125, 0.003125, 0.005, 0.01, 0.02, 0.04, 0.08, 0.16, 0.32\}.$$

This list contains none of the target scales. We compare the fixed bank alone, augmentation by the exact ϵ, duplication of the log-nearest bank scale, and augmentation by 0.75ϵ, 1.25ϵ, 0.5ϵ, or 1.5ϵ. All use B_{64}/H_{32}; H_{52}/H_{52} is the global-family control. The unaugmented bank has 6,704 coefficients, the augmented banks 6,752, and H/H 6,784. Nearest-scale duplication matches the augmented coefficient count without adding function span. The Gaussian initialization scale is $0.005/\sqrt{m_d}$. The four-scale audit is single-seed, with every arm retained in Table 19. Some nominally distinct offsets coincide with a fixed-bank atom: for example, 1.25ϵ equals the nearest bank scale at $\epsilon = 0.0025$ and 0.004. Identical configurations with identical seeds produce identical errors; these rows are not independent replications.

Table 19: A4 scale sensitivity. Relative L_2, seed 42, $R = 16$, 3,000 updates. All eight frozen arms are reported at each diffusion scale; no exact target scale occurs in the fixed bank.

Scale-bank intervention	Params	0.0015	0.0025	0.004	0.008
Exact scale	6,752	1.37×10^{-3}	1.90×10^{-3}	1.97×10^{-3}	5.29×10^{-3}
Fixed bank only	6,704	0.2729	0.0639	7.05×10^{-3}	3.90×10^{-3}
Nearest-scale duplicate	6,752	0.3079	0.0949	7.76×10^{-3}	0.0120
-25% offset	6,752	0.0426	6.69×10^{-3}	8.24×10^{-3}	1.82×10^{-3}
$+25\%$ offset	6,752	0.0163	0.0949	7.76×10^{-3}	0.0120
-50% offset	6,752	0.0440	0.0922	3.99×10^{-3}	2.91×10^{-3}
$+50\%$ offset	6,752	0.1077	0.0786	4.82×10^{-3}	0.0120
H_{52}/H_{52}	6,784	0.8517	0.8696	0.8359	0.4182

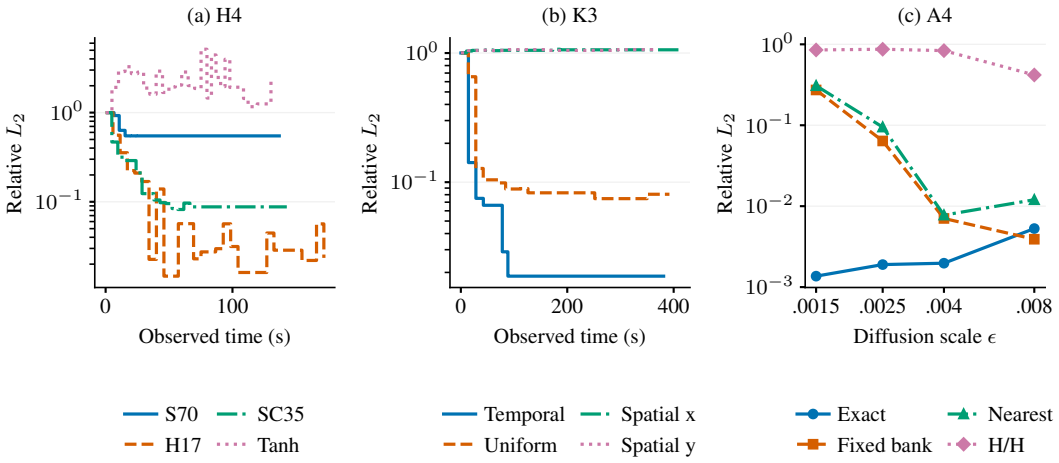


Figure 3: Observed resource trajectories and scale sensitivity, seed 42. Left and middle: reference error of the best-physics checkpoint available at each time, with no reference-based selection or extrapolation. Shared GPU load makes these descriptive trajectories, not an equal-time Pareto comparison. Right: A4 with exact, fixed-bank, duplicated-nearest, and global H/H factors. The full offset study is in Table 19.

Observed error–time trajectories. The plotted error belongs to the lowest-physics-loss checkpoint available at each recorded time. Curves are step functions ending at the last observation. Timing starts after model, optimizer, and validation-set construction and includes training, validation, reference monitoring, and checkpoint-state copying, but not final serialization. Runs use shared RTX 4090 hardware; co-tenant load is uncontrolled. The seed-42 trajectories describe observed costs, rather than an equal-time benchmark; endpoint statistics use all three seeds.

Rank sensitivity with fixed profiles. We separately compare $R \in \{8, 16, 32\}$ on H1, A1, and A4 at seed 42 and 3,000 updates. H1 uses H_{32}/H_{32} factors, its numerical reference, combined physics loss, and cosine learning-rate schedule. A1 uses H_{48}/H_{64} factors, its numerical reference at diffusion 0.01, the common hard lift, and a two-stage learning-rate schedule. Both start at 0.002; A1 switches after 40% of updates to a cosine stage starting at one tenth of that rate. No reference value is used in training. A4 uses the exact-scale augmentation of the present scale study at $\epsilon = 0.0025$, with no change to the factors. All three studies change parameter count with rank and also change the optimization problem, so they cannot identify the irreducible interaction error δ_R in isolation.

Table 20: Rank sensitivity at fixed factors, 3,000 updates, seed 42. Relative L_2 errors are physics-selected. Parameter count grows linearly with rank; this is not a parameter-matched intervention. H1 uses cosine decay, A1 uses a two-stage schedule, and A4 follows the scale-sensitivity protocol.

Case	Fixed factors	$R = 8$	$R = 16$	$R = 32$
H1	H_{32}/H_{32}	0.0136	8.29×10^{-3}	4.32×10^{-3}
A1	H_{48}/H_{64}	0.1621	0.0228	0.0139
A4	B_{64}/H_{32}, exact ϵ	3.32×10^{-3}	1.90×10^{-3}	3.68×10^{-3}

E.1 Fixed-profile CDR evaluation without a manufactured target

On $\Omega = (0, 1)^2$, we specify

$$-0.0075\Delta u + \partial_x u + 0.5u = \sin(\pi y)\big[1 + 0.5\sin(2\pi x y)\big], \qquad u|_{\partial\Omega} = 0. \tag{57}$$

No analytic target is constructed. The source couples coordinates, but this alone does not quantify the solution's separation rank. The known positive x drift motivates localized factors; the test asks whether fixed profiles remain useful, not whether an automatic rule selects them. It is one related CDR problem chosen using physical structure, not a task-population sample.

Frozen configurations and training. The three profiles in Table 5, the scale bank $\widetilde{\mathcal{S}}_0 \cup \{0.64\}$, and training settings are fixed before numerical reference generation or model fitting. The bank excludes 0.0075; the broad 0.64 atom preserves the main matched parameter budgets without inserting a task-specific scale. No family, capacity, atom, or hyperparameter is selected using the new task's errors. Each fit trains fresh coefficients at $R = 16$, with Gaussian initialization scale $0.005/\sqrt{m_d}$, Adam learning rate 0.002, clipping norm 10, and 3,000 updates. The shared loss, sample counts, boundary weight 10, and physics-only checkpoint selection follow the A4 protocol above. Seeds are 42, 2026, and 3407. Validation, evaluation, boundary-validation, and layer-evaluation seeds are 2026091701–2026091704; training interior and boundary seeds are respectively $900000 + s + k$ and $1900000 + s + k$ at update k and run seed s, with edge offsets $10d + b$ for axis $d \in \{0, 1\}$ and side $b \in \{0, 1\}$. All arms share these streams. Runs use CPU float32 with one thread per fit; no speed comparison is made. Reference values are accessed only after all nine fits finish.

Independent reference and evaluation. Float64 centered second-order finite differences on 257^2, 513^2, and 1025^2 grids provide independent reference fields. A discrete sine transform diagonalizes the y Laplacian, followed by tridiagonal x solves; small-grid results agree with an independent sparse solve. The maximum discrete relative residual is 3.05×10^{-12}. Bilinear interpolation evaluates the fields at 16,384 fixed uniform points and 4,096 points in $x > 0.98$. The $257 \to 513$ and $513 \to 1025$ relative reference differences are 0.001513 and 0.000423 globally, and 0.009587 and 0.002406 in the layer. The finest-pair global discrepancy is well below the B/H mean error; grid agreement is not a rigorous continuum-error bound. Global profile orderings are unchanged between the two finest references. All nine results against the finest grid appear in Table 21.

Table 21: All paired seeds for the fixed-profile CDR test. Checkpoints minimize physics validation, never reference error. Boundary RMS is an absolute error.

Seed	Profile	Selected update	Global relative L_2	Layer relative L_2	Boundary RMS
42	B_{64}/H_{32}	2700	0.005051	0.006937	0.002853
42	B_{44}/B_{44}	1400	0.012103	0.007299	0.005069
42	H_{52}/H_{52}	2400	0.304387	0.231177	0.077835
2026	B_{64}/H_{32}	2300	0.006329	0.007419	0.002640
2026	B_{44}/B_{44}	1600	0.010344	0.008435	0.002021
2026	H_{52}/H_{52}	2900	0.283790	0.221657	0.068690
3407	B_{64}/H_{32}	2200	0.007708	0.008194	0.001650
3407	B_{44}/B_{44}	2200	0.007688	0.005904	0.003203
3407	H_{52}/H_{52}	3000	0.270630	0.226833	0.060183

Both B-containing profiles beat H/H at all seeds. B/H has lower global error than B/B at two seeds; their third-seed errors are nearly equal. The small difference in mean layer error is below the seed variation and does not establish a significant advantage for either profile. These results support fixed-profile usefulness on this non-manufactured problem, not cross-family transfer or a representation-only account of trained accuracy.

F Additional Attribution Controls

Table 22 changes one declared design choice per pair, retaining rank, constraints, sampling rules, optimizer settings, and the 3,000-update budget. Profiles and hyperparameters remain fixed across the three seeds. K3 follows Section E; A3/H4 follow their complete-solver protocols, not the matched H4 family study.

Table 22: Targeted attribution controls at 3,000 updates and seeds 42, 2026, 3407. Entries are relative L_2 means $\pm$ sample standard deviations. Parameter counts are identical within each pair. Bold marks the lower mean, not statistical significance. Each arm selects its checkpoint by physics validation.

Case	Params.	Original setting	Control setting	Original error	Control error
K3	2,304	$H_4/H_4/H_{25}$	$H_4/H_9/H_{20}$	0.01499 ± 0.00373	$\mathbf{0.01317 \pm 0.00237}$
A3	1,792	PDE-projected	Random initialization	$\mathbf{0.004014 \pm 0.000188}$	0.24430 ± 0.08926
H4	512	Spectral weighting	Identity weighting	$\mathbf{(1.468 \pm 1.280) \times 10^{-5}}$	0.00946 ± 0.01393

K3: preserving target-frequency coverage. $H_4/H_9/H_{20}$ reallocates capacity while retaining 2,304 parameters and both required temporal atoms, $\sin(20t)$ and $\cos(20t)$. Its errors are 0.014027, 0.014985, and 0.010483 in seed order, versus 0.018713, 0.015005, and 0.011262 for $H_4/H_4/H_{25}$. Both profiles improve on uniform allocation. The original ratio is thus not necessary for the observed benefit. Non-target atoms and dictionary-sized initializations still differ; three seeds establish neither equivalence nor frequency coverage as the sole mechanism.

A3: initialization within a fixed representation. Only the PDE projection in Section A.3 is disabled. The steerable factors, 1,792 coefficients, and Adam rate 0.001 with cosine decay are unchanged. Both arms use CPU execution: four threads for seed 42 and two for the remaining seeds. Random-initialization errors are 0.310080, 0.280125, and 0.142688, compared with 0.004232, 0.003914, and 0.003898 with projection. A3's low error therefore reflects the dictionary and initialization together, not the dictionary alone. This fixed-protocol ablation does not bound the accuracy attainable after separately tuning random initialization.

H4: preconditioning and checkpoint selection. Both arms retain S_{16}/S_{16}, 512 trainable coefficients, constant Adam rate 0.02, and a 32×32 midpoint grid. Let R_{ij} and F_{ij} be the sine projections of the strong residual and forcing. The spectral loss is

$$\mathcal{L}_{\text{spec}}(W) = \frac{\sum_{ij}(W_{ij}R_{ij})^2}{\sum_{ij}(W_{ij}F_{ij})^2}, \qquad W_{ij} = \left[1 - \pi^2(i^2 + j^2)\right]^{-1}.$$

The control sets $W_{ij} = 1$; both retain the additional 0.05-weighted pointwise residual, normalized by the detached mean of operator and forcing squares. Validation uses each arm's own objective, so the intervention affects both optimization and checkpoint selection. Without weighting, errors are 0.025537, 0.001942, and 0.000899. At seed 42, physics validation selects step 200 with error 0.025537; the final step has error 0.002584, and the lowest monitored reference error is 0.001144 at step 2,000. These latter values are diagnostics only: no reference-selected checkpoint replaces the reported result. The ablation quantifies the weighting's effect on the reported solver accuracy, not a pure optimization-speed or representation effect. K3/H4 run on shared GPUs; these internal controls do not equalize external solver priors or compute.

S4: spatial error audit. For the existing checkpoint, the bands $|x - 0.25| \le 0.02$ and $|x - 0.75| \le 0.02$ cover 8.01% of the evaluation grid but contribute only 6.38% of the total squared scalar-flux error. Error is not predominantly localized in these bands; its dominant cause remains unidentified.