SPECTRAL ALLOCATION AND ROUTING FOR PHYSICS-INFORMED LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Physics-informed neural networks (PINNs) often struggle with spatially varying high-frequency solutions of partial differential equations (PDEs) due to spectral bias. Existing approaches, such as Fourier features, periodic activations, multiscale architectures, and adaptive Fourier representations, mitigate this bottleneck but often rely on either globally learned spectral representations or locally adapting predefined spectral bases. We introduce Spectral Allocation and Routing (SAR), an adaptive mechanism that learns a global set of Fourier frequencies and controls their local contributions through continuous, frequency-conditioned spatial routing. This formulation enables direct inspection of the learned spectral content and provides a straightforward way to incorporate spectral inductive biases through structured support parameterizations. Controlled experiments on nonstationary analytical targets investigate spectral discovery and local routing, while neural tangent kernel (NTK) analysis examines their effects on learning dynamics. We further evaluate SAR on heterogeneous Helmholtz problems, where it substantially improves reconstruction accuracy over unrouted representations and achieves low errors across spatially varying and localized multiscale wavefields.

1 INTRODUCTION

Standard multilayer perceptrons (MLPs) typically learn low-frequency components before high-frequency ones, a behavior known as spectral bias (Rahaman et al., 2019; Tancik et al., 2020). This is particularly relevant for physics-informed neural networks (PINNs) (Raissi et al., 2019), where oscillatory solutions and differential operators also introduce optimization difficulties (Wang et al., 2021; 2022). Fourier features and periodic activations can improve high-frequency approximation (Tancik et al., 2020; Sitzmann et al., 2020), but fixed Fourier representations require their frequency range to be chosen in advance.

Existing adaptations address complementary questions: *which frequencies should the model represent, and where should they be used?* Globally trainable frequencies address the first question (Fazliani et al., 2025; Khademi and Dufour, 2025), whereas spatially adaptive methods often adjust the local contributions of predefined frequencies (Hertz et al., 2021; Liu et al., 2025; Shi et al., 2026). Few approaches address both questions simultaneously (Zeraatkar and Tešić, 2026). We propose Spectral Allocation and Routing (SAR), a method to jointly learn which frequencies to represent globally and how strongly each contributes at different spatial locations. Frequencies can be optimized independently for greater flexibility or organized into learnable geometric supports that impose spectral structure. A continuous, frequency-conditioned router then adjusts their local contributions. Both components are optimized jointly during training.

We study the mechanisms separately using analytical signals and projected NTK responses, then evaluate their combination on heterogeneous Helmholtz problems with spatially varying wavefronts and localized oscillations. Our main contributions are:

- Joint spectral allocation and continuous spatial-frequency routing, with optional geometric spectral priors;
- Controlled experiments separating global discovery and local allocation, together with a projected NTK analysis; and

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.

- Helmholtz ablations and comparisons assessing routing, spectral parameterization, and performance across heterogeneous wavefields.

2 Related Work

Spectral bias and adaptive coordinate representations. Fourier features mitigate spectral bias by modifying the effective neural tangent kernel (NTK) (Tancik et al., 2020), while periodic activations provide an alternative for representing oscillatory signals (Sitzmann et al., 2020). More flexible representations include FINER, which adjusts the supported frequency range through variable-periodic activations and bias initialization (Liu et al., 2024), and WIRE, which introduces Gabor wavelet activations localized in space and frequency (Saragadam et al., 2023). These approaches improve the representation of high-frequency content through different feature parameterizations. Other methods explicitly adapt spectral representations to local signal properties. SAPE progressively exposes positional encodings using local optimization feedback (Hertz et al., 2021), while recent approaches introduce spatially adaptive filtering of fixed Fourier features (Shi et al., 2026) or learn coordinate-dependent frequency parameters (Zeraatkar and Tešić, 2026). These methods illustrate complementary strategies for adapting spectral content globally or across spatial regions.

Frequency-aware physics-informed learning. High-frequency approximation is particularly challenging in physics-informed neural networks (PINNs), where the differential operators also affect optimization. Multiscale Fourier features incorporate multiple frequency scales to address NTK-related learning difficulties (Wang et al., 2021). SAFE-NET introduces trainable Fourier-feature frequencies (Fazliani et al., 2025), while TSA-PINN uses sinusoidal activations with trainable neuron-wise frequencies (Khademi and Dufour, 2025). FRES combines adaptive Fourier embeddings with multiscale residual correction (Hou et al., 2026), and IFeF-PINN enhances a learned latent representation with random Fourier features and iterative optimization (Wu et al., 2026). Other methods introduce spatial adaptation: Liu et al. progressively expose predefined Fourier frequencies according to local loss feedback (Liu et al., 2025). For wavefield problems, Alkhalifah and Huang (2024) introduce localized Gabor representations incorporating the governing wavenumber, while Abedi et al. (2026) propose an enhanced Gabor architecture with a learned coordinate transformation. These approaches differ in their use of trainable frequencies, spatial adaptation, and prior knowledge of the governing problem. SAR investigates the combination of freely trainable frequencies and continuous, frequency-conditioned spatial routing, with optional geometric spectral constraints.

Spectral learning dynamics. Beyond representation, spectral bias is closely related to optimization. NTK theory provides a framework for studying neural network learning dynamics (Jacot et al., 2018). Fourier features modify the effective kernel spectrum (Tancik et al., 2020), while NTK analyses of PINNs reveal eigenvector-dependent learning behavior and differences in the convergence rates of loss components (Wang et al., 2021; 2022). Differential operators can introduce additional loss-conditioning difficulties (Krishnapriyan et al., 2021; Rathore et al., 2024), and recent studies further examine the effects of optimization and loss design on spectral learning (Khodakarami et al., 2026). Spatially adaptive Fourier filtering has also been investigated through the NTK (Shi et al., 2026). Motivated by these findings, we examine how global spectral adaptation changes the projected NTK response and relate these changes to reconstruction accuracy and convergence.

3 Method

3.1 Problem setting

Let $\mathbf{x} \in \Omega \subset \mathbb{R}^d$ denote spatial coordinates and $u(\mathbf{x})$ the target solution. We study spectral adaptation in two settings: supervised approximation of analytical functions with known frequency content and physics-informed learning of solutions satisfying

$$\mathcal{N}[u](\mathbf{x}) = f(\mathbf{x}), \qquad \mathcal{B}[u](\mathbf{x}) = g(\mathbf{x}), \tag{1}$$

where $\mathcal{N}$ is a differential operator and $\mathcal{B}$ defines the boundary conditions. In this setting, the model is trained by minimizing

$$\mathcal{L}_{\text{PINN}} = \lambda_r \mathcal{L}_r + \lambda_b \mathcal{L}_b + \lambda_d \mathcal{L}_d, \tag{2}$$

where $\mathcal{L}_r$, $\mathcal{L}_b$, and $\mathcal{L}_d$ denote the PDE residual, boundary-condition, and optional data losses, respectively.

The decoder, spectral support, and router are jointly optimized, with automatic differentiation applied through the entire representation to evaluate the PDE residual. No separate frequency estimation or spectral supervision is required.

3.2 Spectral Allocation and Routing

SAR combines trainable Fourier frequencies with continuous, frequency-conditioned spatial routing. Spectral allocation determines which frequencies are represented globally, while routing controls their contributions across the spatial domain. Both components are optimized jointly.

Spectral allocation. We represent the global spectral support using M wavevectors $\mathcal{K} = \{\mathbf{k}_m\}_{m=1}^{M}$, with $\mathbf{k}_m \in \mathbb{R}^d$. Each defines sine and cosine features $\boldsymbol{\phi}_m(\mathbf{x}) = [\sin(2\pi\mathbf{k}_m^\top \bar{\mathbf{x}}), \cos(2\pi\mathbf{k}_m^\top \bar{\mathbf{x}})]^\top$, where $\bar{\mathbf{x}}$ denotes normalized coordinates.

In the unconstrained formulation, the wavevectors are independent trainable parameters, allowing their frequencies and orientations to change during optimization. Alternatively, geometric supports organize groups of frequencies within a shared spectral structure. In two dimensions, each elliptical support is parameterized as

$$\mathbf{k}_{s,j} = \mathbf{c}_s + R(\theta_s)\,\mathrm{diag}(r_{s,1}, r_{s,2})\mathbf{q}_j, \tag{3}$$

where $\mathbf{c}_s$ is the support center, θ_s its orientation, $r_{s,1}, r_{s,2}$ its radii, and $\mathbf{q}_j$ are fixed offsets within the unit disk. This reduces the number of independent spectral parameters and introduces a geometric prior, although the resulting constraints may limit adaptation when the relevant frequencies do not follow the assumed structure.

Continuous spatial routing. The routing function $\alpha_m(\mathbf{x}) = r_\psi(\bar{\mathbf{x}}, \bar{\mathbf{k}}_m)$ controls the local contribution of each mode using its position and normalized wavevector $\bar{\mathbf{k}}_m = \mathbf{k}_m \oslash \mathbf{f}_{\max}$, where $\mathbf{f}_{\max}$ defines the frequency scale. In the analytical experiment E1B from Section 4.2, shallow networks produce either frequency-conditioned logits or the center and radius of a local Gaussian spectral mask. Both use softmax gating across modes, although their parameterizations differ (Appendix A.1).

For the Helmholtz experiments, we instead use Gaussian radial basis functions (RBFs) in frequency space. This factorizes the routing computation into spatial and spectral components, avoiding repeated neural network evaluations for every position-frequency pair. A spatial network produces $\mathbf{g}_\psi(\bar{\mathbf{x}}) \in \mathbb{R}^B$, while B fixed Gaussian RBFs encode each normalized wavevector through $h_b(\bar{\mathbf{k}}) = \exp(-\|\bar{\mathbf{k}} - \boldsymbol{\nu}_b\|^2/(2\sigma_f^2))$, where $\boldsymbol{\nu}_b$ are the RBF centers and σ_f controls their width. Both representations are combined as

$$r_\psi(\bar{\mathbf{x}}, \bar{\mathbf{k}}) = 2\,\mathrm{sigmoid}\left(\frac{\mathbf{g}_\psi(\bar{\mathbf{x}})^\top \mathbf{h}(\bar{\mathbf{k}})}{\sqrt{B}}\right). \tag{4}$$

This construction provides smooth routing over space and frequency, allowing nearby wavevectors to share spectral basis functions. Unlike the E1B softmax router, it assigns independent weights in $(0, 2)$, allowing multiple frequencies to be emphasized simultaneously. Zero initialization of its final layer produces unit weights, preserving the original Fourier representation at initialization.

The resulting representation is

$$\mathbf{z}(\mathbf{x}) = \begin{bmatrix} \boldsymbol{\alpha}(\mathbf{x}) \odot \sin(2\pi\mathcal{K}\bar{\mathbf{x}}) \\ \boldsymbol{\alpha}(\mathbf{x}) \odot \cos(2\pi\mathcal{K}\bar{\mathbf{x}}) \end{bmatrix}, \qquad u_\Theta(\mathbf{x}) = h_\theta(\mathbf{z}(\mathbf{x})), \tag{5}$$

where $\boldsymbol{\alpha}(\mathbf{x})$ collects the routing weights and $\mathcal{K}$ contains the wavevectors as rows.

Implementation settings for the geometric supports and RBF router are provided in Appendix A.2.

3.3 Effect on learning dynamics

Spectral placement also influences learning dynamics. For model u_Θ, the empirical NTK is $K_\Theta(\xi, \xi') = \nabla_\Theta u_\Theta(\xi)^\top \nabla_\Theta u_\Theta(\xi')$. Using an orthonormal spectral basis Q, we examine

$$H_\Theta = Q^* \mathbf{K}_\Theta Q, \qquad \rho_j = [H_\Theta]_{jj}, \tag{6}$$

where ρ_j measures the projected kernel response along a spectral direction. It represents a kernel eigenvalue only when Q diagonalizes the kernel; off-diagonal terms otherwise couple spectral directions. For the fixed-frequency linear readout in E1A, the kernel is stationary and its spectral response depends directly on support placement. We examine the response during global adaptation without assuming independent mode-wise convergence for the nonstationary routed model. Derivations and empirical verification are provided in Appendix A.3.

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

We consider three spectral representations: *Fixed* uses predefined Fourier frequencies, *Free* learns each frequency independently, and *Geometric* learns structured spectral supports. Each is evaluated with and without routing, giving six configurations. Reconstruction accuracy is measured by relative L^2 error, $\varepsilon_{L^2} = \|\hat{u} - u^*\|_2 / \|u^*\|_2$. The Helmholtz experiments use five seeds; implementation and evaluation details are given in Appendix A.

4.2 E1: CONTROLLED SPECTRAL ADAPTATION

We first investigate the support and routing capabilities of SAR using one-dimensional signals with known frequency content, trained directly on samples without physics-informed losses. With a linear readout, the representation is

$$\hat{u}(x) = c + \sum_{m=1}^{M} r(x, k_m)[a_m \cos(2\pi k_m x) + b_m \sin(2\pi k_m x)], \tag{7}$$

with $r(x, k_m) = 1$ when routing is disabled. Both experiments use the target

$$u^*(x) = 0.35 \sin(4\pi x) + \exp\left(-\frac{x^2}{2\sigma^2}\right) \cos\big(2\pi(k_0 x + \alpha x^2)\big)\,, \tag{8}$$

where $x \in [-1, 1]$, $k_0 = 36$, $\alpha = 10$, and σ controls the spatial extent of the oscillatory component. By varying σ, we examine global spectral discovery and local adaptation separately.

E1A: Global spectral discovery. Our first experiment investigates whether structured spectral supports can discover a high-frequency component when their center is initialized away from its dominant frequency region. We set $\sigma = 0.18$, concentrating the oscillatory component near the origin and limiting the range of instantaneous frequencies with substantial signal amplitude. This allows us to focus on global spectral discovery.

We compare fixed Fourier grids, independently trainable frequencies, and a one-dimensional structured support $k_m = \mu + Rs_m$, with fixed relative positions s_m. The center and radius can be learned jointly, or the radius progressively annealed; we also include a case with the ideal frequency grid (oracle). All models use a linear readout without routing.

Figure 1 shows that both annealing R and joint optimization allow the support to shift and contract toward the relevant frequency band, sharply reducing reconstruction error to a level comparable with the oracle grid. Fixed wide grids and independently trainable frequencies achieve lower accuracy under the tested conditions. These observations illustrate how coordinated support adaptation can facilitate spectral discovery.

The projected NTK responses in Fig. 2 provide an additional view of the learning dynamics. As the support adapts, its projected kernel response becomes concentrated near the target frequency band, approaching the oracle response. This evolution accompanies the improvement in reconstruction accuracy, while the calculated and automatic-differentiation responses show close agreement.

E1B: Local spectral allocation. We next investigate whether spatial-frequency routing can adapt a fixed Fourier representation to a signal whose frequency changes with position. We increase the localization width to $\sigma = 0.35$, exposing the varying instantaneous frequency $f_{\text{inst}}(x) = 36 + 20x$ over a larger region and providing a more demanding test of local adaptation.

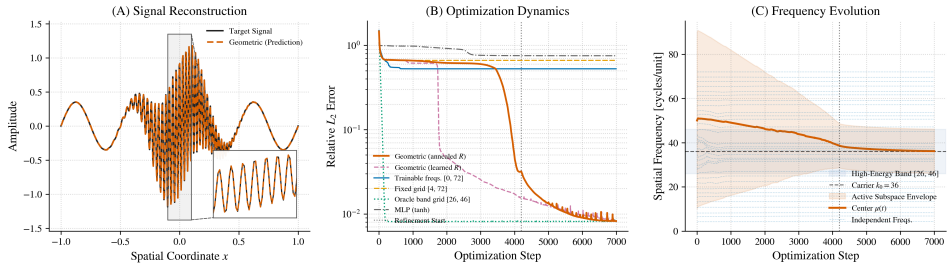


Figure 1: **E1A: Global spectral discovery.** (A) Reconstruction of the localized signal. (B) Relative L^2 error of the spectral baselines and radius variants. (C) Evolution of the support center and active band versus independent frequencies and the target carrier. The vertical dotted line marks the end of support contraction.

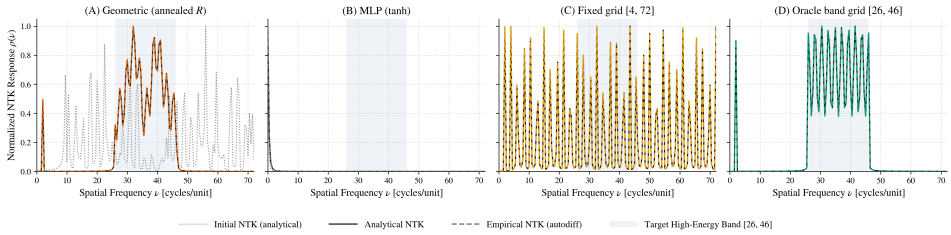


Figure 2: **Projected NTK response in E1A.** Normalized responses for (A) the structured representation, (B) a $\tanh$ MLP, (C) a fixed wide grid, and (D) an oracle band grid. Dotted, solid, and dashed curves show initial calculated, final calculated, and final automatic-differentiation responses, respectively.

The routed models share a fixed bank of 32 frequencies covering $[16, 56]$. They are implemented through shallow MLPs with one hidden layer of eight neurons. The generic router predicts a weight for each position-frequency pair, while the structured router predicts the center $c(x)$ and radius $\rho(x)$ of a Gaussian spectral mask. Both apply softmax across modes. The structured router therefore learns a translated and dilated spectral window rather than independent routing weights. We additionally compare unrouted fixed and globally adaptive representations and coordinate-dependent frequency variants. Unlike these alternatives, the router approaches change the spatial contributions of a fixed frequency bank without modifying its frequencies. No explicit frequency supervision is used.

Figure 3 shows that the structured Gaussian mask achieves reconstruction error below 10^{-3}, improving upon the generic router and unrouted models. Its learned center closely follows the analytical instantaneous frequency within the oscillatory region, while the radius determines the width of the local selection window. These results suggest that the structured Gaussian profile provides a useful inductive bias for local spectral selection. Since the global frequency bank remains fixed, E1B isolates local allocation, complementing the global spectral discovery investigated in E1A.

4.3 E2: Heterogeneous Helmholtz problems

We next evaluate SAR under physics-informed training on two-dimensional heterogeneous Helmholtz problems on $\Omega = [-1, 1]^2$. We consider two complementary benchmarks: phase-warped waves with spatially varying frequency and direction, and incident waves combined with localized radial oscillations. Both use manufactured solutions: we prescribe an exact solution and spatially varying refractive index, then compute the corresponding source term. These benchmarks test the representation of controlled wave patterns rather than physical scattering simulations.

Governing equation and training. We consider

$$\Delta u(\mathbf{x}) + \kappa_0^2 n^2(\mathbf{x}) u(\mathbf{x}) = q(\mathbf{x}), \quad \mathbf{x} \in \Omega, \tag{9}$$

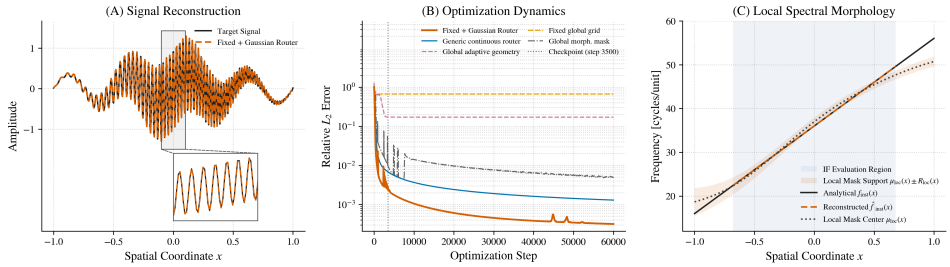


Figure 3: **E1B: Local spectral allocation.** (A) Reconstruction with Gaussian routing. (B) Mean squared reconstruction error throughout optimization, with a marker at step 3,500. (C) Learned mask center and support versus analytical and reconstructed instantaneous frequencies; shading marks the diagnostic interval.

where κ_0 is the reference wavenumber and $n(\mathbf{x})$ the spatially varying refractive index. We impose Dirichlet conditions $u = u^*$ on $\partial\Omega$, with $q = \Delta u^* + \kappa_0^2 n^2 u^*$. Models minimize normalized PDE residual and boundary losses without interior solution data. All use 23,000 interior and 3,000 boundary points and a common 30,000-step Adam protocol with initial learning rate 10^{-2} and cosine decay. Additional settings are detailed in Appendix A.2.

GRF phase-warped waves. Our first benchmark uses Gaussian random fields (GRFs) to introduce smooth spatial changes in wave frequency and direction:

$$u^*(\mathbf{x}) = \sin\left[2\pi f_0 \left(\mathbf{d}^\top \mathbf{z} + \alpha_\phi \ell \psi_\phi(\mathbf{z})\right)\right], \tag{10}$$

where $\mathbf{z}$ denotes normalized coordinates, $\mathbf{d}$ the incident direction, ψ_ϕ a random phase field, α_ϕ its modulation strength, and ℓ its correlation length. An independent random field defines the refractive-index variations. We use $f_0 = 10$ and $\ell = 0.15$, with fixed random realizations across methods (Appendix A.2).

Localized multiscale particle waves. The second benchmark combines an incident plane wave with P localized radial oscillations:

$$u^*(\mathbf{x}) = u_{\text{inc}}(\mathbf{z}) + \frac{A}{\sqrt{P}} \sum_{j=1}^{P} c_j \exp\left[-\left(\frac{\rho_j(\mathbf{z})}{\delta_j}\right)^2\right] \sin(2\pi f_j \rho_j(\mathbf{z})), \tag{11}$$

where u_{inc} is the incident wave, ρ_j the regularized radial distance from particle j, and c_j, δ_j, and f_j determine its amplitude, spatial extent, and oscillation frequency. The refractive index includes smooth localized variations around the particles. We set $f_{\text{inc}} = 8$ and $A = 0.40$, and evaluate configurations with four and eight particles. The latter retains the first four particles and introduces additional oscillatory components. Since increasing P also changes the refractive index and per-component amplitude, this experiment examines combined spatial and spectral complexity rather than particle count alone. Exact particle parameters and remaining definitions are given in Appendix A.2.

4.4 E2A: Architectural Ablations

We compare the six configurations on the GRF and four-particle problems under identical initial frequency banks, mode counts, decoders, and training protocols; parameter counts differ with spectral parameterization and routing. Table 1 reports final errors, and Fig. 4 shows convergence and representative reconstructions. Without routing, none achieves a mean error below 0.7 on either benchmark. Routing substantially improves all three: even *Fixed + Router* reaches 0.0805 and 0.131 on GRF and particles, while *Free + Router* achieves the lowest means, 0.0574 and 0.0263, reductions of approximately 14-fold and 28-fold relative to unrouted *Free*.

Geometric + Router also improves substantially but does not outperform *Free + Router*. A possible explanation is that a shared geometric support restricts independent frequency placement when the relevant spectrum does not follow its imposed shape; the ablation alone does not establish this mechanism.

Table 1: **Architectural ablations.** Final relative L^2 error (mean $\pm$ sample standard deviation across five seeds). The lowest mean error in each column is highlighted in bold.

Architecture	Parameters	GRF $f_0 = 10$	Particles $P = 4$
Fixed	41,217	0.893 ± 0.0133	0.785 ± 0.0226
Free	41,729	0.792 ± 0.0325	0.744 ± 0.0519
Geometric	41,237	0.86 ± 0.0337	0.77 ± 0.0796
Fixed + Router	43,489	0.0805 ± 0.00536	0.131 ± 0.00992
Free + Router	44,001	$\mathbf{0.0574 \pm 0.0054}$	$\mathbf{0.0263 \pm 0.00212}$
Geometric + Router	43,509	0.0653 ± 0.0149	0.0502 ± 0.0209

Table 2: **External comparisons.** Final relative L^2 error (mean $\pm$ sample standard deviation over five seeds). Carrier-informed and frequency-blind Gabors are separate configurations. Best mean result in each column is highlighted in bold.

Method	GRF $f_0 = 10$	Particles $P = 4$	Particles $P = 8$
SAR (Free + Router)	0.0574 ± 0.0054	$\mathbf{0.0263 \pm 0.00212}$	$\mathbf{0.024 \pm 0.00197}$
MFF	0.713 ± 0.109	0.777 ± 0.0746	–
SAFE-Fourier	0.14 ± 0.0227	0.18 ± 0.0264	–
Gabor-wavefields (blind)	0.538 ± 0.178	0.56 ± 0.0531	0.628 ± 0.0778
Gabor-wavefields (informed)	0.499 ± 0.174	0.492 ± 0.0949	0.565 ± 0.0966
Gabor-enhanced (blind)	0.166 ± 0.03	0.185 ± 0.105	0.264 ± 0.131
Gabor-enhanced (informed)	$\mathbf{0.0258 \pm 0.00214}$	0.0757 ± 0.0387	0.225 ± 0.296

The representative *Free + Router* solutions reproduce the main GRF wavefront variations and both the incident wave and localized particle oscillations (Fig. 4).

4.5 E2B: Comparison with existing methods

We compare SAR with multiscale Fourier features (MFF) (Wang et al., 2021), a Fourier-only adaptation of SAFE-NET (SAFE-Fourier) (Fazliani et al., 2025), and two adapted Gabor architectures: Gabor-wavefields (Alkhalifah and Huang, 2024) and Gabor-enhanced (Abedi et al., 2026). Both Gabor architectures are evaluated with frequency-blind and carrier-informed initialization, yielding six baseline configurations. All methods share the same physics-informed objective and main Adam training schedule. Implementation details are provided in Appendix C.

We evaluate both Gabor architectures under two initialization conditions. Frequency-blind variants use SAR's nominal fixed frequency bank, whereas carrier-informed variants receive the known incident or carrier frequency, but not the individual particle frequencies. However, note that matching the nominal bank does not guarantee identical effective wavevectors because each architecture parameterizes its spectral representation differently. Gabor-wavefields models also receive the known refractive-index field under both conditions.

Table 2 shows that SAR achieves lower mean reconstruction error than all evaluated frequency-blind baselines on each applicable benchmark. Carrier-informed Gabor-enhanced achieves a lower mean error on GRF (0.0258 versus 0.0574), illustrating the potential benefit of prior spectral information. On the four- and eight-particle problems, SAR achieves the lowest reported mean errors (0.0263 and 0.0240, respectively), including comparisons with the carrier-informed Gabor architectures. These results demonstrate the effectiveness of joint spectral adaptation and spatial routing on the evaluated localized multiscale problems, while highlighting the influence of prior frequency information on the comparisons.

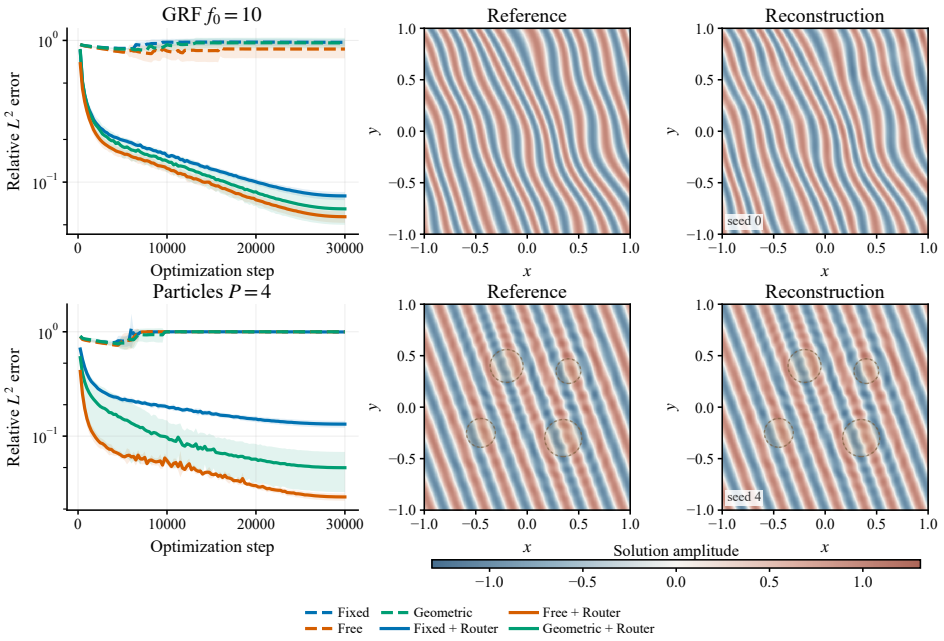


Figure 4: **Architectural ablations.** GRF (top) and four-particle (bottom) convergence and representative reconstructions. Solid curves denote routed models and dashed curves unrouted models; lines and shading show the mean and one sample standard deviation over five seeds. Solutions are shown on a shared signed-amplitude scale, with particle regions indicated below.

4.6 E2C: Interpretation of Learned Spectral Allocation

Beyond reconstruction accuracy, SAR allows us to inspect its learned frequencies and routing weights. Figure 5 shows that *Free + Router* learns a nonuniform wavevector distribution on the four-particle benchmark, with the strongest spatially averaged router emphasis near the incident carrier and lower average weights assigned to other frequency groups. These weights indicate relative spectral emphasis, not physical Fourier energy or the importance of a mode at a particular location. A frequency with a low spatial average may still be strongly activated within a localized region.

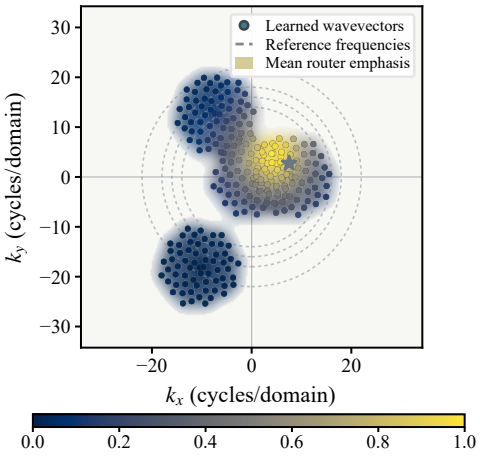


Figure 5: **Learned spectral allocation.** SAR wavevectors colored by mean routing weights. Dashed contours indicate reference frequency regions, the star marks the carrier, and the heatmap interpolates router emphasis.

5 Discussion and Limitations

Our experiments highlight the complementary roles of spectral adaptation and spatial-frequency routing. The controlled experiments show that structured spectral supports can discover relevant frequency regions and that local routing can follow spatially varying frequency content. The Helmholtz ablations further show that routing provides the largest improvement in reconstruction accuracy under the tested conditions. In particular, independently trainable frequencies offer limited improvements without routing, whereas combining both mechanisms achieves the lowest mean errors in the two benchmarks. Geometric constraints also improve substan-

tially when combined with routing, although they do not outperform freely trainable frequencies in these experiments.

The external comparisons highlight the ability of SAR to adapt with limited spectral prior knowledge. Without predefining the carrier frequency, *Free + Router* achieves consistently low reconstruction errors across all three Helmholtz benchmarks, including the more complex eight-particle configuration. Although carrier-informed Gabor-enhanced achieves a lower mean error on the GRF problem, its performance deteriorates and becomes more variable on the particle benchmarks. In contrast, SAR maintains low errors and limited variability across all three problems. Inspection of its learned spectral representation also reveals nonuniform frequency placement and selective routing, the spatially averaged routing weights in the particle problem focus around the carrier.

Several limitations remain. First, our Helmholtz benchmarks use manufactured solutions with fixed problem configurations. Although they allow controlled comparisons, they do not establish generalization to arbitrary heterogeneous media or physical scattering problems. Second, routing introduces additional parameters and computational overhead, so the computational efficiency of the method cannot be completely isolated despite the available indicators. Moreover, the external methods differ in architecture, frequency information, and method-specific optimization procedures, which had to be adapted to our benchmark problems. Nevertheless, under a common optimization protocol, SAR achieves lower errors than the frequency-blind baselines without requiring additional optimization procedures. Finally, spectral allocation is not unique: the learned frequencies and routing weights provide insight into the representation but should not be interpreted directly as a recovered physical spectrum. Broader evaluations across problem realizations, training budgets, and PDE families are needed to establish the generality of these findings.

6 Conclusion

We introduced Spectral Allocation and Routing (SAR), a mechanism that jointly learns global Fourier frequencies and their spatially varying contributions to improve physics-informed learning. Controlled experiments demonstrate global spectral discovery and local frequency allocation, while heterogeneous Helmholtz benchmarks show substantial improvements from routing, particularly when combined with independently trainable frequencies. SAR achieves consistently low reconstruction errors on both spatially varying wavefields and problems with multiple localized oscillations, maintaining its performance as spatial and spectral complexity increases. Overall, these results support adaptive spectral allocation and routing as complementary mechanisms for representing solutions with spatially varying frequency content.

Reproducibility statement

The analytical experiments, benchmark definitions, data-generation procedures, architectural configurations, and training and evaluation protocols are described in the main text and appendix. The supplementary material provides additional implementation details, additional spectral diagnostics, computational-cost measurements, and adaptations of the external baselines. Anonymized source code and experiment configurations will be released to facilitate reproduction of the reported results.

AI use statement

Generative AI tools were used to provide feedback on research methodology and experimental design, assist with the development and refinement of experimental code, and support the interpretation of experimental results. They were also used for literature search and summarization, identification of relevant prior work, manuscript structuring, and drafting and editing of preliminary text. All AI-assisted suggestions, code, analyses, and literature claims were reviewed and verified by the authors, who take full responsibility for the scientific claims, experimental results, and final content of this work.

References

Mohammad Mahdi Abedi, David Pardo, and Tariq Alkhalifah. Gabor-enhanced physics-informed neural networks for fast simulations of acoustic wavefields. *Neural Networks*, 193:107978, 2026.

doi: 10.1016/j.neunet.2025.107978.

Tariq Alkhalifah and Xinquan Huang. Physics-informed neural wavefields with gabor basis functions. *Neural Networks*, 175:106286, 2024. doi: 10.1016/j.neunet.2024.106286.

Shaghayegh Fazliani, Zachary Frangella, and Madeleine Udell. Enhancing physics-informed neural networks through feature engineering. *Transactions on Machine Learning Research*, 2025. SAFE-NET.

Amir Hertz, Or Perel, Raja Giryes, Olga Sorkine-Hornung, and Daniel Cohen-Or. Sape: Spatially-adaptive progressive encoding for neural optimization. In *Advances in Neural Information Processing Systems*, volume 34, pages 8820–8832, 2021.

Bo-Ya Hou, Yu-Long Bai, Xia-Ting Jing, and Chun-lin Huang. Fourier feature-enhanced multi-layer residual stacking network: A novel multiscale modeling approach for physics-informed neural networks. *Neural Networks*, 195:108247, 2026. doi: 10.1016/j.neunet.2025.108247.

Arthur Jacot, Franck Gabriel, and Clément Hongler. Neural tangent kernel: Convergence and generalization in neural networks. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

Amirhossein Khademi and Steven Dufour. Physics-informed neural networks with trainable sinusoidal activation functions for approximating the solutions of the navier–stokes equations. *Computer Physics Communications*, 314:109672, 2025. doi: 10.1016/j.cpc.2025.109672.

Siavash Khodakarami, Vivek Oommen, Nazanin Ahmadi Daryakenari, Maxim Beekenkamp, and George Em Karniadakis. Spectral bias in physics-informed and operator learning: Analysis and mitigation guidelines. *Computer Methods in Applied Mechanics and Engineering*, 461:119156, 2026. doi: 10.1016/j.cma.2026.119156.

Aditi S. Krishnapriyan, Amir Gholami, Shandian Zhe, Robert M. Kirby, and Michael W. Mahoney. Characterizing possible failure modes in physics-informed neural networks. In *Advances in Neural Information Processing Systems*, volume 34, pages 26548–26560, 2021.

Yarong Liu, Hong Gu, Xiangjun Yu, and Pan Qin. Diminishing spectral bias in physics-informed neural networks using spatially-adaptive fourier feature encoding. *Neural Networks*, 182:106886, 2025. doi: 10.1016/j.neunet.2024.106886.

Zhen Liu, Hao Zhu, Qi Zhang, Jingde Fu, Weibing Deng, Zhan Ma, Yanwen Guo, and Xun Cao. Finer: Flexible spectral-bias tuning in implicit neural representation by variable-periodic activation functions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2713–2722, 2024.

Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 5301–5310. PMLR, 2019.

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

Pratik Rathore, Weimu Lei, Zachary Frangella, Lu Lu, and Madeleine Udell. Challenges in training PINNs: A loss landscape perspective. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 42159–42191. PMLR, 2024.

Vishwanath Saragadam, Daniel LeJeune, Jasper Tan, Guha Balakrishnan, Ashok Veeraraghavan, and Richard G. Baraniuk. Wire: Wavelet implicit neural representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18507–18516, June 2023.

Ligen Shi, Jun Qiu, Yuhang Zheng, and Chang Liu. Adaptive local frequency filtering for fourier-encoded implicit neural representations. *arXiv preprint arXiv:2604.02846*, 2026.

Vincent Sitzmann, Julien N. P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In *Advances in Neural Information Processing Systems*, 2020.

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

Sifan Wang, Hanwen Wang, and Paris Perdikaris. On the eigenvector bias of fourier feature networks: From regression to solving multi-scale pdes with physics-informed neural networks. *Computer Methods in Applied Mechanics and Engineering*, 384:113938, 2021. doi: 10.1016/j.cma.2021.113938.

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

Yulun Wu, Miguel Aguiar, Karl H. Johansson, and Matthieu Barreau. Iterative training of physics-informed neural networks with fourier-enhanced features. In *International Conference on Learning Representations*, 2026. ICLR 2026.

Ehsan Zeraatkar and Jelena Tešić. Safir: Spatially-aware activation function for implicit neural representations. *Computer Vision and Image Understanding*, 268:104754, 2026. doi: 10.1016/j.cviu.2026.104754.

A Experimental Details

A.1 Controlled spectral experiments

Both analytical experiments use 512 uniformly spaced samples on $[-1, 1]$ and minimize the mean squared reconstruction error. They include an explicit low-frequency Fourier component at two cycles per unit, allowing the remaining spectral capacity to focus on the high-frequency component. Reconstruction errors are evaluated on the same sampling grid used during training; these experiments test spectral adaptation rather than generalization to unseen samples.

E1A: Global spectral discovery. All Fourier models use 32 high-frequency modes, an explicit two-feature low-frequency branch, and a linear readout without bias. We additionally evaluate a single-hidden-layer $\tanh$ MLP with 32 neurons. Models are trained for 7,000 iterations using Adam with a learning rate of 3×10^{-3} and three initialization seeds $\{41, 42, 43\}$. Representative results correspond to the run with median final relative L^2 error.

The fixed Fourier grid spans $[4, 72]$, while the oracle grid spans $[26, 46]$ cycles per unit. The independently trainable frequencies are initialized uniformly over $[4, 72]$ and constrained to $(0, 72)$ through a sigmoid parameterization.

The geometric representation defines its frequencies as

$$k_m^{(j)} = \mu^{(j)} + R(j)s_m, \qquad s_m = -1 + \frac{2(m-1)}{M-1}, \tag{12}$$

where j denotes the optimization iteration and $M = 32$. The support center is trainable and initialized at $\mu = 50$. Its radius contracts linearly from 40 to 10 during the first 60% of training, followed by 40% of refinement at the final radius.

We additionally evaluate a geometric variant that jointly optimizes the center and radius, both initialized at 50 and 40, respectively, without scheduled contraction. A separate spectral-window variant learns a normalized Gaussian-shaped weighting over a fixed Fourier dictionary and includes a

radius penalty of $10^{-3}R$. All remaining models minimize reconstruction MSE without additional regularization.

These configurations distinguish independent frequency learning, jointly optimized spectral geometry, spectral weighting, and coordinated support movement with scheduled contraction.

E1B: Local spectral allocation. All models use 32 high-frequency modes, an explicit two-feature low-frequency branch, and a linear readout without bias. We use a Gaussian envelope with $\sigma = 0.35$ and a chirp with instantaneous frequency $f_{\text{inst}}(x) = 36 + 20x$.

The fixed global frequency bank is centered at 36 cycles per unit with radius 20, spanning $[16, 56]$. The fixed-grid baseline and both locally routed models share this bank. A separate global-adaptation baseline jointly learns the center and radius without spatial routing, allowing local frequency selection to be examined independently of global spectral adaptation.

The generic continuous router uses a shared $2 \to 8 \to 1$ network with $\tanh$ activations. It predicts routing logits from spatial position and normalized frequency, followed by softmax normalization across modes.

The structured local mask instead uses a $1 \to 8 \to 2$ network to predict only a position-dependent center $c(x)$ and radius $\rho(x)$. Its routing weights are defined as

$$\alpha_m(x) = \frac{\exp\left[-\frac{1}{2}\left(\frac{s_m - c(x)}{\rho(x)}\right)^2\right]}{\sum_{j=1}^{M} \exp\left[-\frac{1}{2}\left(\frac{s_j - c(x)}{\rho(x)}\right)^2\right]}. \tag{13}$$

The normalized center is bounded to $(-1, 1)$, while the radius is constrained to be positive with a minimum value of 0.02. They are initialized at $c(x) = 0$ and $\rho(x) = 1$ throughout the domain. Both routers use softmax normalization, so their weights sum to one across modes. The generic router initially assigns uniform weights of $1/M$, whereas the structured mask initially produces a Gaussian-shaped distribution centered at zero. In contrast, the unrouted baselines use unit feature weights. Consequently, the routed and unrouted representations differ in their initial feature scaling, although all models learn their linear readout coefficients. The comparison evaluates the complete implemented representations rather than isolating routing under identical initialization.

We additionally consider two coordinate-dependent frequency representations. The first directly predicts individual local frequencies, while the second predicts a local spectral center and radius from which the frequencies are constructed. Neither uses explicit routing weights.

All models are trained for 60,000 iterations using Adam with AMSGrad and a constant learning rate of 3×10^{-3}. The reported experiment uses one initialization seed. Reconstruction error is also recorded at iteration 3,500.

The reconstructed instantaneous frequency is estimated post-training from the phase derivative of the complex high-frequency representation. Diagnostics are restricted to positions where the target envelope exceeds or equals 0.15, with squared-envelope weighting. The analytical instantaneous frequency is never used for training.

A.2 Additional Helmholtz implementation details

Geometric supports and RBF router. For the Helmholtz experiments, SAR uses 256 Fourier modes and four spectral supports. The *Geometric* representation learns the center, orientation, and two radii of each support, while *Free* optimizes individual wavevector components independently.

The spatial-frequency router uses 32 Gaussian RBFs distributed over normalized frequency space, with width $\sigma_f = 0.35$. Its spatial network contains one hidden layer with 64 $\tanh$ neurons. These settings are shared across the routed configurations.

Manufactured source and losses. With u^* and n prescribed, the forcing is $q(\mathbf{x}) = \Delta u^*(\mathbf{x}) + \kappa_0^2 n^2(\mathbf{x}) u^*(\mathbf{x})$. With $\widetilde{\mathcal{R}}_\Theta$ denoting the residual normalized by the source root-mean-square magni-

Table 3: **Particle benchmark parameters.** Centers are given in physical coordinates. Radii, frequencies, and decay parameters follow the normalized-domain construction.

Center	Radius	Contrast	Frequency	Decay
$(-0.45, -0.25)$	0.07	0.5	14	0.28
$(0.35, -0.30)$	0.09	0.8	18	0.25
$(-0.20, 0.40)$	0.08	0.6	22	0.30
$(0.40, 0.35)$	0.06	0.9	16	0.24
$(-0.50, 0.30)$	0.07	0.7	24	0.26
$(0.05, -0.05)$	0.08	0.5	20	0.30
$(0.55, 0.00)$	0.06	0.8	26	0.23
$(0.10, 0.55)$	0.07	0.6	17	0.27

tude, the evaluated loss is

$$\mathcal{L} = \frac{1}{N_r} \sum_{i=1}^{N_r} |\widetilde{\mathcal{R}}_\Theta(\mathbf{x}_i)|^2 + \frac{1}{N_b} \sum_{j=1}^{N_b} |u_\Theta(\mathbf{x}_j) - u^*(\mathbf{x}_j)|^2. \tag{14}$$

GRF refractive index. The independently sampled index field is $n(\mathbf{x}) = 1 + \alpha_n \tanh[\psi_n(\mathbf{z})]$, with fixed realizations across models.

Random-field generation. The GRF benchmark approximates its random fields using 64 random Fourier components:

$$\psi(\mathbf{z}) = \sqrt{\frac{2}{J}} \sum_{j=1}^{J} a_j \cos(2\pi \mathbf{q}_j^\top \mathbf{z} + b_j), \tag{15}$$

where $J = 64$ and

$$a_j \sim \mathcal{N}(0, 1), \qquad b_j \sim \mathcal{U}(0, 2\pi), \qquad \mathbf{q}_j \sim \mathcal{N}\left(\mathbf{0}, \frac{\mathbf{I}}{(2\pi\ell)^2}\right). \tag{16}$$

Independent random realizations are generated for the phase and refractive-index fields, using seeds 0 and 100, respectively. Both fields remain fixed across architectures and training seeds.

Particle wave and medium definitions. The regularized radial coordinate and incident wave are

$$\rho_j(\mathbf{z}) = \sqrt{\|\mathbf{z} - \mathbf{z}_j\|^2 + \epsilon^2}, \qquad u_{\text{inc}}(\mathbf{z}) = \sin(2\pi f_{\text{inc}} \mathbf{d}^\top \mathbf{z}), \tag{17}$$

and the smooth index profile is

$$n(\mathbf{x}) = 1 + \sum_{j=1}^{P} c_j \,\text{sigmoid}\left(\frac{L r_j - \|\mathbf{x} - \mathbf{x}_j\|}{L\tau}\right), \tag{18}$$

where L is the domain width, r_j the normalized particle radius, and τ the transition width. We use $\tau = 0.015$ and $\epsilon = 0.01$.

Particle configurations. Table 3 provides the particle parameters omitted from the main text. The four-particle benchmark uses the first four rows, while the eight-particle benchmark uses all eight.

Checkpoint selection and evaluation. During training, the combined physics-informed loss is evaluated every 250 steps using a fixed subset of 4,096 interior collocation points and all boundary points. The checkpoint with the lowest recorded loss is retained. This subset is drawn from the training collocation points and is not an independent validation set.

The PDE residual is normalized by the root-mean-square magnitude of the source term, computed once on a 101×101 reference grid. Intermediate convergence curves use a separate 101×101

solution grid, while final reconstruction errors are evaluated on a 401×401 grid after restoring the selected checkpoint.

The training seeds change model initialization, interior collocation points, and minibatch sampling. For each benchmark and seed, the different architectures receive the same collocation points. The underlying reference solution remains fixed.

These choices allow comparisons across architectures while separating training variability from changes in the reference problem.

A.3 Spectral learning dynamics and NTK analysis

This section details the NTK analysis used in E1A. We first characterize the effect of spectral placement on the readout kernel and then extend the analysis to the complete finite-width models. Finally, we describe the numerical Fourier projection and the independent verification using automatic differentiation.

Fourier readout kernel. Consider the one-dimensional Fourier representation used in E1A, consisting of an explicit low-frequency component, M high-frequency modes, and a linear readout without bias:

$$\begin{aligned} u_\Theta(x) = a_0 \sin(4\pi x) + b_0 \cos(4\pi x) \\ + \sum_{m=1}^{M} \left[a_m \sin(2\pi k_m x) + b_m \cos(2\pi k_m x) \right]. \end{aligned} \tag{19}$$

The high-frequency support is parameterized as

$$k_m = \mu + R s_m, \qquad s_m = -1 + \frac{2(m-1)}{M-1}, \tag{20}$$

where μ and R denote the spectral center and radius, respectively.

To isolate the effect of spectral placement, we first consider the NTK associated only with the linear readout parameters, treating all frequencies as fixed.

The readout Jacobian contains the sine and cosine features themselves. Consequently, the corresponding kernel is

$$\begin{aligned} K_{\mathrm{ro}}(x, x') = \sin(4\pi x)\sin(4\pi x') + \cos(4\pi x)\cos(4\pi x') \\ + \sum_{m=1}^{M} \left[\sin(2\pi k_m x)\sin(2\pi k_m x') + \cos(2\pi k_m x)\cos(2\pi k_m x') \right]. \end{aligned} \tag{21}$$

Using the trigonometric addition identity, this simplifies to

$$K_{\mathrm{ro}}(x, x') = \cos\left(4\pi(x - x')\right) + \sum_{m=1}^{M} \cos\left(2\pi k_m (x - x')\right). \tag{22}$$

The first term corresponds to the explicit low-frequency component. The second term contains the high-frequency contribution defined by the high-frequency Fourier modes.

For fixed global frequencies, this kernel is stationary because it depends only on $x - x'$. Its spectral measure consists of symmetric components at the represented positive and negative frequencies. In particular, the high-frequency contribution has the formal spectral representation

$$\mathcal{F}\{K_{\mathrm{ro}}^{\mu,R}\}(\nu) = \frac{1}{2} \sum_{m=1}^{M} \left[\delta(\nu - k_m) + \delta(\nu + k_m) \right], \tag{23}$$

where the low-frequency contribution has been omitted for clarity.

This expression shows explicitly how changing the spectral center or radius changes the frequencies represented by the readout kernel. However, it does not characterize the complete NTK when the spectral parameters are also trainable.

Complete NTK of the geometric model. The main E1A model learns its spectral center while contracting its radius according to an externally prescribed schedule. Its frequencies at optimization iteration j are

$$k_m^{(j)} = \mu(\vartheta) + R(j)s_m, \qquad \mu(\vartheta) = k_{\max}\,\mathrm{sigmoid}(\vartheta), \tag{24}$$

where $k_{\max} = 72$ and ϑ is the trainable center parameter.

The radius is initialized at 40 and decreases linearly to 10 during the first 4,200 iterations. It is subsequently held constant for the remaining 2,800 iterations.

Because the radius is prescribed rather than optimized, it contributes no trainable parameter derivatives to the NTK. Nevertheless, it changes the frequencies and therefore the kernel evaluated at each iteration.

Let Θ_{ro} denote the readout parameters. The complete trainable parameter vector is

$$\Theta = (\Theta_{\mathrm{ro}}, \vartheta). \tag{25}$$

Differentiating Eq. (19) with respect to the spectral center gives

$$\frac{\partial u_\Theta(x)}{\partial \mu} = 2\pi x \sum_{m=1}^{M} \left[a_m \cos(2\pi k_m x) - b_m \sin(2\pi k_m x)\right]. \tag{26}$$

Applying the chain rule to the sigmoid parameterization yields

$$\frac{\partial \mu}{\partial \vartheta} = k_{\max}\,\mathrm{sigmoid}(\vartheta)\left[1 - \mathrm{sigmoid}(\vartheta)\right]. \tag{27}$$

Defining

$$g_\mu(x) = \frac{\partial u_\Theta(x)}{\partial \vartheta} = \frac{\partial u_\Theta(x)}{\partial \mu}\frac{\partial \mu}{\partial \vartheta}, \tag{28}$$

the complete NTK of the geometric model becomes

$$K_\Theta^{\mathrm{geo}}(x, x') = K_{\mathrm{ro}}^{\mu,R}(x, x') + g_\mu(x) g_\mu(x'). \tag{29}$$

The first term describes the contribution of the Fourier readout. The second term accounts for the trainable spectral center.

Importantly, the second term is generally nonstationary because it depends separately on x and x'. Therefore, although the readout kernel is stationary at any fixed spectral configuration, the complete geometric NTK need not be stationary.

Furthermore, the prescribed radius schedule changes the representation explicitly during optimization. The evolution of the model therefore depends on both parameter updates and the externally scheduled spectral contraction. A single fixed-kernel analysis does not fully characterize this training trajectory.

NTKs of the comparison models. We compute the complete NTK for the four architectures presented in Fig. 2.

For the fixed-grid and oracle-grid models, the Fourier frequencies are constant and only the readout coefficients are trainable. Their complete NTKs therefore coincide with Eq. (22). The fixed grid spans $[4, 72]$, whereas the oracle grid spans $[26, 46]$ cycles per unit.

The MLP baseline consists of a single hidden layer with 32 $\tanh$ neurons:

$$u_\Theta^{\mathrm{MLP}}(x) = \sum_{h=1}^{32} v_h \tanh(w_h x + b_h) + c. \tag{30}$$

Writing $z_h(x) = w_h x + b_h$, its exact NTK is

$$\begin{aligned} K_\Theta^{\mathrm{MLP}}(x, x') = 1 &+ \sum_{h=1}^{32} \tanh(z_h(x)) \tanh(z_h(x')) \\ &+ \sum_{h=1}^{32} v_h^2 \,\mathrm{sech}^2(z_h(x))\,\mathrm{sech}^2(z_h(x'))(xx' + 1). \end{aligned} \tag{31}$$

This expression includes derivatives with respect to the input weights, hidden biases, output weights, and output bias.

Unlike the fixed Fourier models, the MLP kernel is generally nonstationary. We therefore analyze all four architectures using the same numerical spectral projection, without assuming that Fourier modes diagonalize their complete NTKs.

Numerical Fourier projection. We evaluate the complete NTK on a periodic analysis grid containing $N = 320$ points,

$$x_i = -1 + \frac{2i}{N}, \qquad i = 0, \dots, N-1. \tag{32}$$

This grid is distinct from the 512-point reconstruction grid used for training. Its half-open interval avoids duplicating the periodic endpoint.

Let $\mathbf{K}_\Theta$ denote the $N \times N$ matrix obtained by evaluating the complete NTK on this grid.

We consider the discrete Fourier directions

$$q_\nu(x_i) = \frac{1}{\sqrt{N}} \exp(2\pi \mathrm{i} \nu x_i), \tag{33}$$

with frequencies

$$\nu \in \{0, 0.5, 1, \dots, 72\}. \tag{34}$$

These frequencies correspond to harmonics of the periodic interval of length two. They remain below the Nyquist limit of the analysis grid.

For each frequency, we calculate the projected NTK response

$$\rho_\Theta(\nu) = \mathbf{q}_\nu^* \left(\frac{\mathbf{K}_\Theta}{N} \right) \mathbf{q}_\nu. \tag{35}$$

Equivalently, collecting the Fourier directions in an orthonormal matrix Q gives

$$H_\Theta = Q^* \left(\frac{\mathbf{K}_\Theta}{N} \right) Q. \tag{36}$$

The diagonal entries measure the kernel response along individual Fourier directions, while the off-diagonal entries capture interactions between them.

If the Fourier basis diagonalizes the kernel, these responses coincide with its eigenvalues. Otherwise, they are Rayleigh quotients and must not be interpreted as independent frequency-wise convergence rates.

In particular, stationarity of a kernel on the continuous domain does not by itself guarantee exact diagonalization by the discrete Fourier basis on a finite interval with nonperiodic frequency components.

Relationship with learning dynamics. For a quadratic reconstruction loss

$$\mathcal{L}(\Theta) = \frac{1}{N} \sum_{i=1}^{N} \left[u_\Theta(x_i) - u^*(x_i) \right]^2, \tag{37}$$

let $\mathbf{e}$ denote the vector of prediction errors on the evaluation grid.

Under gradient flow, and in the absence of externally scheduled changes to the representation, the prediction-error dynamics satisfy

$$\frac{d\mathbf{e}}{dt} = -2 \frac{\mathbf{K}_{\Theta(t)}}{N} \mathbf{e}. \tag{38}$$

If the kernel remains constant and the Fourier basis diagonalizes it, the error components along its eigenvectors decay exponentially, with rates determined by the corresponding kernel eigenvalues.

For the trainable geometric model, however, the full NTK changes during optimization. In addition, the prescribed radius schedule explicitly modifies the representation. Consequently, Eq. (38) alone does not describe its complete training dynamics.

We therefore use the projected response as a diagnostic of spectral sensitivity, rather than as a direct measurement of independent frequency-wise convergence rates. Its evolution is examined alongside the reconstruction errors and learned spectral trajectories.

Analytical and automatic-differentiation verification. For numerical evaluation, we construct the analytical Jacobian of each model, including every trainable parameter, and compute its NTK as

$$\mathbf{K}_{\text{analytical}} = JJ^{\top}. \tag{39}$$

Independently, we calculate the Jacobian of the same model using automatic differentiation and construct

$$\mathbf{K}_{\text{autodiff}} = J_{\text{autodiff}} J_{\text{autodiff}}^{\top}. \tag{40}$$

Both calculations use the same model parameters and spectral configuration. The computations are performed in double precision on the CPU.

We verify that the analytical and automatic-differentiation Jacobians contain the same number of trainable parameters and compare the resulting kernel matrices using

$$\varepsilon_K = \frac{\|\mathbf{K}_{\text{analytical}} - \mathbf{K}_{\text{autodiff}}\|_F}{\|\mathbf{K}_{\text{autodiff}}\|_F}. \tag{41}$$

The implementation requires the relative matrix discrepancy to remain below 10^{-8} at every evaluated snapshot.

This independent calculation verifies the analytical finite-width NTK used to obtain the projected responses. It does not establish that the Fourier directions are eigenvectors of the complete kernel.

Training snapshots and visualization. To relate the NTK analysis to the reconstruction experiment, we use the same trained models and representative runs presented in E1A. For each architecture, the representative run is the one with median final relative L^2 error among the three training seeds.

We evaluate the NTK at three stages: initialization, the beginning of refinement after 4,200 parameter updates, and the end of training after 7,000 updates. All three snapshots are retained for analysis.

Figure 2 displays the initial and final projected responses. For visualization, each initial analytical response is normalized by its own maximum. The final analytical and automatic-differentiation responses share a common normalization, given by the maximum of the corresponding final analytical response.

The shaded interval $[26, 46]$ indicates the oracle frequency band used as a visual reference.

These normalized responses allow us to examine changes in the distribution of spectral sensitivity and compare the analytical and numerical calculations. Because normalization is performed separately for each model and training stage, the figure does not provide absolute comparisons of NTK magnitude across architectures or between initialization and the final snapshot.

Together with the reconstruction experiment, this analysis illustrates how coordinated spectral adaptation can alter the frequency distribution of the complete NTK during learning, while distinguishing these observations from theoretical guarantees about mode-wise convergence.

B Additional Experimental Results

B.1 Fourier-domain reconstruction

The spatial reconstructions presented in the main paper are complemented by a frequency-domain analysis. The spectra are displayed on logarithmic scales and normalized using the maximum magnitude of the corresponding reference spectrum. The coefficient-error maps show the Fourier transform of the prediction error, rather than the relative error of individual Fourier coefficients.

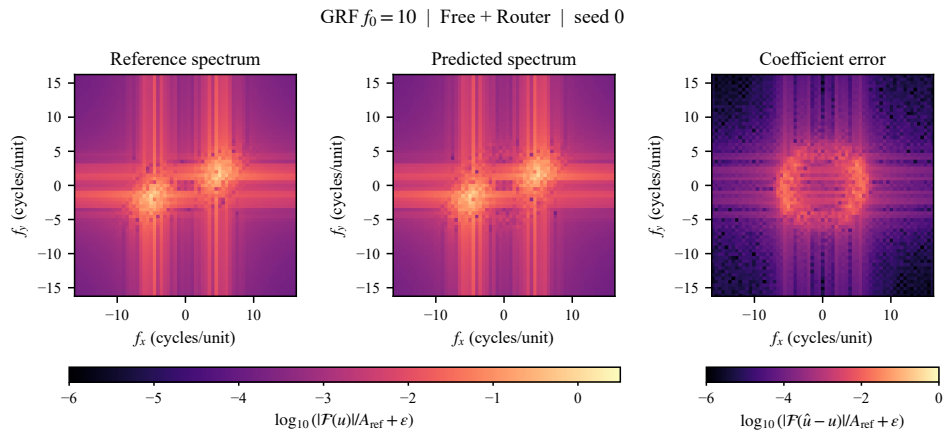


Figure 6: **Fourier-domain reconstruction of the GRF benchmark.** Reference spectrum, predicted spectrum, and Fourier-coefficient error for SAR (Free + Router), seed 0. Reference and predicted magnitudes share the same logarithmic color scale. The error map shows the magnitude of the Fourier transform of the prediction error, normalized by the maximum reference-spectrum magnitude.

GRF benchmark. Figure 6 shows that SAR reproduces the main frequency components of the GRF reference solution, including its two dominant spectral regions and their surrounding structure. The remaining coefficient errors are distributed around these regions and across lower-amplitude spectral components.

This provides a frequency-domain view of the reconstruction shown in the main paper. The residual spectrum indicates where approximation errors remain, rather than establishing that the learned wavevectors individually recover the analytical local frequencies.

Localized particle benchmarks. Figures 7 and 8 show the corresponding spectra for the four- and eight-particle problems. Both reference spectra contain prominent components associated with the incident wave and broader spectral patterns introduced by the localized oscillations.

SAR reproduces the dominant spectral structures in both configurations. The remaining errors are more broadly distributed around the central frequency region, including components associated with the localized oscillations. This complements the spatial reconstructions and provides an additional view of the approximation errors in the particle benchmarks.

Because each spectrum is normalized by its own reference maximum, differences in the displayed coefficient-error maps should not be interpreted directly as differences in total reconstruction error between the two configurations. The relative L^2 errors reported in the main text provide that quantitative comparison.

B.2 Eight-particle spatial reconstruction

Figure 9 complements the eight-particle spectral analysis with the corresponding spatial reconstruction. Together, the spatial and frequency-domain figures provide two views of SAR's approximation on this additional benchmark.

B.3 Computational cost

The increased flexibility of SAR introduces additional computational cost. Table 4 reports instantiated parameter counts and recorded training times, complementing the accuracy comparisons in the main paper.

On the GRF benchmark, mean training time increases from approximately 341 seconds for Fixed to 528 seconds for SAR. The corresponding times on the four-particle problem are 624 and 820 seconds,

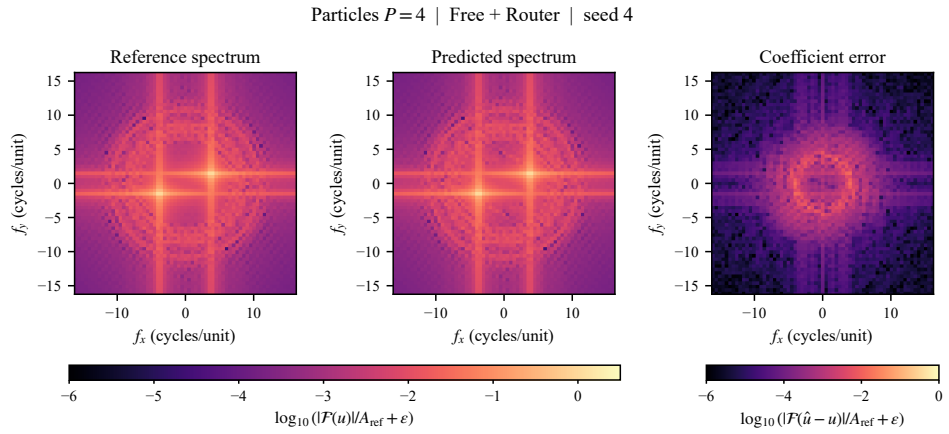


Figure 7: **Fourier-domain reconstruction of the four-particle benchmark.** Reference spectrum, predicted spectrum, and Fourier-coefficient error for SAR (Free + Router), seed 1. The localized oscillatory components introduce spectral structure beyond the dominant incident-wave frequencies. All magnitudes are normalized by the maximum reference-spectrum magnitude.

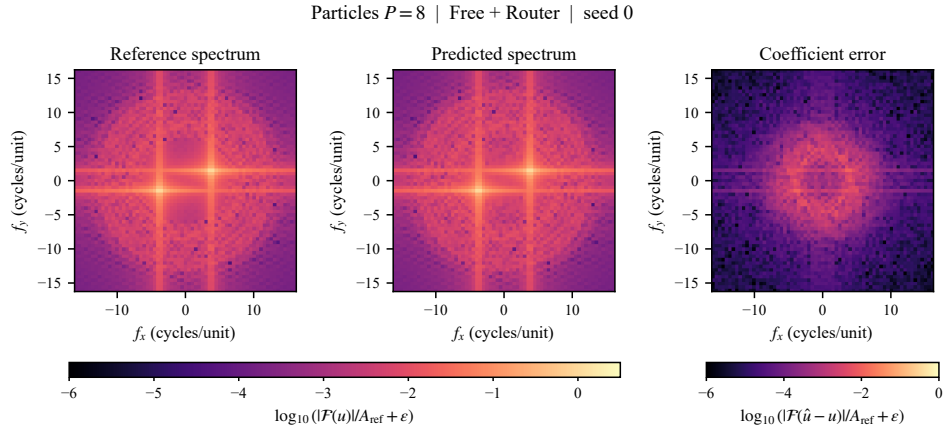


Figure 8: **Fourier-domain reconstruction of the eight-particle benchmark.** Reference spectrum, predicted spectrum, and Fourier-coefficient error for SAR (Free + Router), seed 0. The additional localized oscillations modify the spectral distribution, while the prediction retains the dominant reference components. The normalization follows Fig. 7.

respectively. Thus, the accuracy improvements obtained by combining trainable frequencies and routing come with additional training overhead.

Among the external methods, carrier-informed Gabor-enhanced achieves a lower mean error than SAR on the GRF benchmark but requires approximately 724 seconds, compared with 528 seconds for SAR. These observations illustrate that approximation accuracy and computational cost must be considered separately.

The reported times include periodic training diagnostics and exclude model construction, final high-resolution evaluation, and subsequent figure generation. The values characterize the evaluated implementations rather than hardware-independent computational complexity.

C Implementation of External Baselines

This section documents the architectural and optimization details of the adapted methods, including differences from their original implementations.

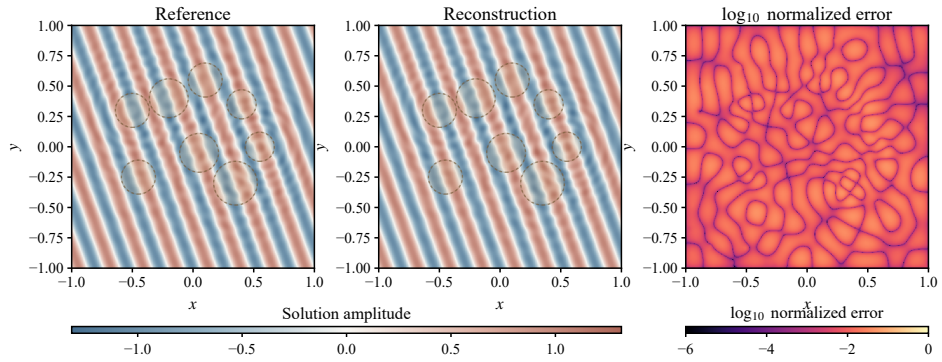


Figure 9: **Eight-particle reconstruction.** Reference solution, SAR reconstruction, and logarithmic normalized absolute error for the eight-particle Helmholtz problem. The reference and reconstruction share the same signed-amplitude scale, and the marked circles indicate the scattering regions.

Table 4: **Recorded training cost.** Mean ± sample standard deviation over available seeds. The lowest mean training time in each column is highlighted in bold. Wall-clock time includes periodic logging and method-specific procedures. Instantiated Gabor-enhanced parameter counts include an unused imaginary output head; effective counts are smaller by the head size.

Method	Inst. params	GRF time (s)	$P = 4$ time (s)	$P = 8$ time (s)
Fixed	41,217	$\mathbf{340.5 \pm 9.123}$	$\mathbf{624 \pm 8.295}$	–
Free	41,729	357.7 ± 8.624	643.1 ± 9.323	–
Geometric	41,237	387.2 ± 8.911	674.4 ± 5.334	–
Fixed + Router	43,489	493.8 ± 11.4	783.8 ± 10.29	–
Free + Router	44,001	527.7 ± 11.72	820.4 ± 10.47	$\mathbf{1126 \pm 41.15}$
Geometric + Router	43,509	559.2 ± 12.87	852.8 ± 8.184	–
MFF	43,489	543.4 ± 19.92	819.8 ± 17.23	–
SAFE-Fourier	33,665	675.3 ± 24.89	945.4 ± 21.19	–
Gabor-wavefields (blind)	44,820	1045 ± 37.68	1681 ± 24.25	2227 ± 69.47
Gabor-wavefields (informed)	44,820	1045 ± 37.69	1683 ± 20.77	2226 ± 69.2
Gabor-enhanced (blind)	39,680	723.7 ± 25.32	1029 ± 10.12	1254 ± 40.62
Gabor-enhanced (informed)	39,680	724.4 ± 23.29	1027 ± 13.41	1268 ± 21.37

Multiscale Fourier features (MFF). We use two fixed Gaussian frequency banks with 128 wavevectors each. Their standard deviations are 1 and 16 cycles over the normalized domain. Each bank is processed by the same three-hidden-layer network, with 96 $\tanh$ neurons per layer. The branch outputs are concatenated and passed through a linear prediction layer.

Unlike the original training configuration, we use the common Adam optimizer and cosine learning-rate schedule. The frozen frequency banks are not optimized during training.

SAFE-Fourier. We retain SAFE-NET's trainable tensor-product harmonics, learnable feature coefficients, per-sample feature normalization, and shallow decoder. The implementation uses 128 harmonic frequency pairs, yielding four tensor-product features per pair, and one hidden layer with 64 $\tanh$ neurons.

For our domain, the initial normalized-domain frequencies are $k_j = j/4$, with $j = 1, \ldots, 128$. The feature coefficients are initialized to one. Before entering the decoder, the features are centered and normalized, using a numerical regularization constant of 10^{-3}.

The original domain-knowledge features are omitted, and optimization follows the common Adam schedule instead of the original Adam/L-BFGS procedure. SAFE-Fourier therefore evaluates the adapted Fourier-feature architecture rather than the complete original SAFE-NET implementation.

Gabor-based architectures. We evaluate adapted implementations of the Gabor-wavefields and Gabor-enhanced architectures.

Gabor-wavefields uses a three-layer sinusoidal network with 96 neurons per layer to predict the Gabor coefficients. An additional network predicts the spatial centers of the Gabor functions. The phase also depends on the known refractive-index field, which is provided to both initialization variants.

Gabor-enhanced uses a positional encoding with four frequency bands and two sinusoidal hidden layers with 64 neurons each. Its Gabor representation includes trainable orientations and velocity parameters, together with learned coordinate-dependent offsets. The Gabor width is fixed to three.

Both architectures are adapted to predict the real-valued benchmark solutions. Only their real output channels contribute to the training objective. The enhanced architecture additionally omits the original problem-specific absorbing-boundary and source-neighborhood treatments, using the common PDE and boundary losses instead.

Gabor initialization. The frequency-blind variants initialize their nominal spectral parameters using SAR's fixed frequency bank. For Gabor-wavefields, this determines the initial orientations and phase-frequency magnitudes. For Gabor-enhanced, the initial orientations and velocity parameters are chosen to reproduce the nominal frequency magnitudes. The zero-frequency mode is handled separately to avoid division by zero.

The carrier-informed variants instead use the known nominal carrier frequency: $f_0 = 10$ for the GRF problem and $f_{\mathrm{inc}} = 8$ for the particle problems. They do not receive the individual particle frequencies or locations through initialization.

These initialization conditions match nominal frequency information but do not produce identical initial representations. In particular, the effective local phase gradients of Gabor-enhanced also depend on its learned coordinate transformation. Moreover, Gabor-wavefields receives the known refractive-index field in both initialization conditions.

For parameter reporting, the enhanced implementation instantiates both real and imaginary output heads, although only the real head participates in training. Its instantiated parameter count is 39,680, while its effective trainable parameter count is 39,424.