Adaptive multiscale operator correction via learned spectral subspace and physics-informed optimization

Subham Patel[1,2] **Himanshu Pandey**[1] **Ratikanta Behera**[1*]

[1]Department of Computational and Data Sciences, Indian Institute of Science, Bengaluru, IN 560012
[2]IISc Mathematics Initiative, Department of Mathematics, Indian Institute of Science, Bengaluru, IN 560012
subhampatel@iisc.ac.in phimanshu@iisc.ac.in ratikanta@iisc.ac.in

Abstract

Neural operators have shown strong performance in learning solution mappings for differential equations and offer fast inference across parameter spaces. However, they often suffer from spectral bias, leading to poor resolution of high-frequency and localized features, especially in out-of-distribution settings. In this work, we propose adaptive multiscale operator correction (AMOC), a hybrid framework that combines neural operator learning with physics-constrained optimization in a low-dimensional, instance-specific adaptive spectral subspace. First, a neural operator learns a coarse solution, which is then projected onto a sparse multiscale basis using either wavelet family or Fourier modes, where dominant components are selected based on energy. Finally, a physics-informed optimization is performed on adaptive spectral subspace, significantly reducing computational cost while improving accuracy. The proposed approach enables efficient correction of neural operator predictions by leveraging the sparsity and localization of multiscale representations, while bypassing automatic differentiation in the residual loss through closed-form derivatives of the chosen basis. The proposed AMOC is evaluated on the heat, Poisson, Darcy, and Helmholtz equations, achieving up to four orders-of-magnitude error reduction over Fourier neural operator, with particularly strong gains observed in challenging out-of-distribution regimes.

1 Introduction

Partial differential equations (PDEs) form the fundamental mathematical basis for modeling processes such as transport, diffusion, wave propagation, and nonlinear interactions across many scientific and engineering fields, including fluid dynamics, electromagnetism, materials science, climate modeling, and biology [1, 2]. Even though traditional numerical methods for solving PDEs are very accurate, they often become too expensive when applied to high-resolution simulations and many-query parameter evaluations [3, 4].

Neural operators have emerged as a powerful alternative by learning mappings directly between infinite-dimensional function spaces, enabling mesh-independent, near-instantaneous parametric inference once trained. DeepONet [5] grounds this approach in a universal approximation theorem for operators, introducing a branch-trunk architecture that efficiently learns nonlinear operators from limited data. The Fourier neural operator (FNO) [6] parameterizes integral kernels in the spectral domain, achieving resolution-invariant learning and strong performance across a wide range of benchmark PDEs. A range of subsequent works have extended these ideas, including multiwavelet-based operator learning [7], factorized Fourier operators [8], and attention-based Galerkin operators

*Corresponding author

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

[9]. In contrast, physics-informed neural networks (PINNs) [10] offer a complementary approach by encoding PDE residuals directly into the training loss, enabling solutions without labeled data, but are limited to a single instance. Despite their success, both classes of methods share a fundamental limitation. They exhibit spectral bias [11, 12], systematically underrepresenting high-frequency components and spatially localized structures, and struggle to resolve sharp transitions or fine-scale features when trained on coarse or low-resolution data. For operator learning, this limitation is especially pronounced in the out-of-distribution regime [13].

To address the limitation, several multiscale PINNs and operators have been proposed. Wang et al. [14] proposed a multi-magnitude loss term to capture the multiscale structure of the PDE solution. Furthermore, wavelet-based PINNs [15, 16] were proposed to better capture localized and multiscale features of PDE solutions using formulation in which the solution is represented in a wavelet basis, allowing for spatial adaptivity and improved representation of sharp gradients. Physics-informed operator methods such as PINO [17] combine the parametric efficiency of neural operators with PDE residual penalties, enabling instance-wise refinement at test time. However, this refinement still does not explicitly enforce multiscale or localized structure. As a result, its ability to resolve highly localized or sharply varying features remains dependent on the expressivity of the FNO training. As a result, despite instance-wise refinement, the method inherits the spectral bias of its underlying representation, and thus achieves limited improvement on high-frequency or localized features.

Inspired by multiscale PINNs, we propose adaptive multiscale operator correction (AMOC), a three-stage framework that addresses these limitations by constructing an instance-specific sparse subspace from neural operator predictions and perform spectral aware physics-constrained optimization. Unlike global fine-tuning approaches, optimization is restricted to a low-dimensional, energy-ranked spectral subspace with closed-form derivative. This per-instance refinement trades inference speed for accuracy, yielding improvements of several orders of magnitude over the FNO baseline, with the largest gains observed in out-of-distribution regimes where data-driven generalization degrades.

The main contributions of our work are as follows:

1. We propose AMOC, a hybrid framework that combines neural operator inference with instance-specific physics-constrained optimization in a learned low-dimensional spectral subspace.
2. We introduce an adaptive spectral subspace construction strategy that selects dominant basis functions from neural operator predictions via energy-based sparsification.
3. We demonstrate that optimization to an instance-adaptive spectral subspace leads to substantial improvements in accuracy, particularly in out-of-distribution regimes with high-frequency or localized features.
4. We apply AMOC with both wavelet-based (W-AMOC) and Fourier-based (F-AMOC) variants, providing a controlled study of localized versus global spectral representations for PDE correction.

2 Methodology

2.1 Problem setup

Let $\Omega \subset \mathbb{R}^d$ be a bounded domain. We consider a class of parametric PDEs of the form:

$$\begin{cases} \mathcal{N}[u;\varepsilon] = f, & \mathbf{x} \in \Omega, \\ \mathcal{B}[u] = g, & \mathbf{x} \in \partial\Omega, \end{cases} \tag{1}$$

where $u : \Omega \to \mathbb{R}$ is the unknown solution field, $\varepsilon \in \mathcal{E}$ is a parametric input field, $\mathcal{N}$ is a differential operator, and $\mathcal{B}$ encodes boundary or initial conditions. The task is to learn the solution operator $\mathcal{G} : \mathcal{E} \to \mathcal{U}$ and then enhance the solution accuracy via spectral aware physics-constrained optimization.

2.2 Overview of the proposed architecture

Neural operators such as FNO and DeepONet approximate solution operators efficiently, but exhibit spectral bias, systematically underrepresenting high-frequency and localized structures. We demonstrate an adaptive multiscale operator correction framework to address the limitations of both

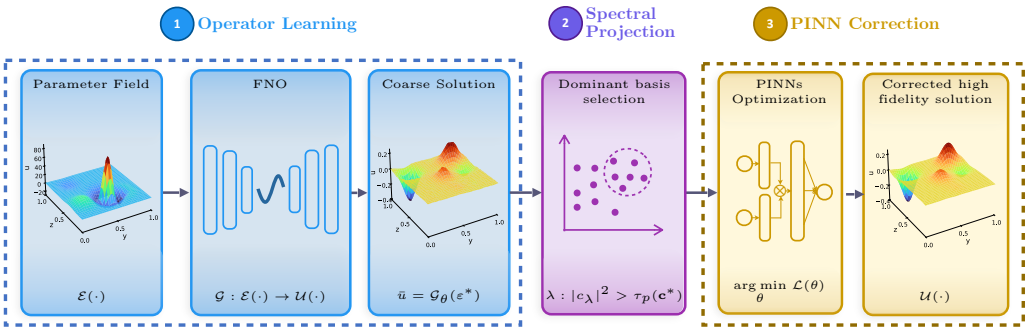

Figure 1: Schematic diagram of the AMOC model. From left to right: Stage I-operator learning, where a coarse solution is obtained, Stage II-dominant basis selection via spectral projection, Stage III-physics-constrained correction via spectral PINNs to obtain the final high-fidelity solution.

data-driven operator learning and instance-specific physics-informed methods. The method constructs a sparse subspace $\mathcal{V}^*$ during the learned neural operator inference and performs physics-constrained optimization within it. This yields significant improvements in accuracy by multiple orders of magnitude, particularly for out-of-distribution regimes.

The AMOC architecture consists of three stages, as shown in Fig. 1. In Stage I, a neural operator produces a coarse prediction $\bar{u} = \mathcal{G}_\theta(\varepsilon)$. In Stage II, this prediction is represented in a sparse spectral subspace $\mathcal{V}^*$, from which a compact set of parameters is obtained. In Stage III, the solution is parameterized within $\mathcal{V}^*$ and constrained to satisfy the governing PDE, yielding the final high-fidelity solution u_{θ^*}.

The expressive power of Stage II is governed by the choice of spectral representation. We consider two variants, one based on spatially localized multiscale functions and another based on global spectral modes. These share identical formulations in Stages I and III and differ only in the representation used in Stage II, enabling a controlled comparison of localized versus global features. The detailed construction of each is provided below.

2.3 Stage I: Neural operator approximation

We train a neural operator $\mathcal{G}_\theta : \mathcal{E} \to \mathcal{U}$ from a dataset $\{(\varepsilon^{(i)}, u^{(i)})\}_{i=1}^N$ of input-output pairs by minimizing the data loss

$$\mathcal{L}_{\text{data}}(\theta) = \frac{1}{N}\sum_{i=1}^{N} \left\| \mathcal{G}_\theta(\varepsilon^{(i)}) - u^{(i)} \right\|^2_{\mathcal{L}_2(\Omega)}.$$

In our implementation, we employ the FNO, which operates by lifting the input to a latent channel space, applying a sequence of Fourier integral operators in the spectral domain, and projecting back to the solution space. For a new instance ε^*, inference is immediate:

$$\bar{u} = \mathcal{G}_\theta(\varepsilon^*).$$

This prediction serves as the input to Stage II. Any sufficiently expressive neural operator can be employed in this stage.

2.4 Stage II: Adaptive spectral subspace projection

The goal of this stage is to identify a compact spectral subspace that captures the dominant basis of the coarse prediction while discarding redundant ones. This enables subsequent physics-constrained modeling to operate in a significantly reduced subspace. While the coarse prediction $\bar{u}$ captures the dominant low-frequency structure, it fails to resolve fine-scale and localized features due to spectral truncation. We therefore represent $\bar{u}$ in a structured multiscale spectral basis that captures these localized variations.

Wavelet dictionary. To this end, we construct a family of Gaussian wavelets in d-dimensions, parameterized by dyadic scale indices $\mathbf{j} = (j_1, \ldots, j_d)$ and translation indices $\mathbf{k} \in \mathbb{Z}^d$:

$$\psi_{\mathbf{j},\mathbf{k}}(\mathbf{x}) = \left(\prod_{i=1}^{d} \sqrt{j_i}\right) \left(\prod_{i=1}^{d} (-(j_i x_i - k_i))\right) \exp\left(-\frac{1}{2}\sum_{i=1}^{d} (j_i x_i - k_i)^2\right).$$

To account for the non-compact support of the Gaussian envelope, we introduce a padding parameter $a > 0$ and define admissible translations as

$$k_i \in \left[\lfloor (x_i^{\min} - a)\, j_i \rfloor,\ \lceil (x_i^{\max} + a)\, j_i \rceil\right].$$

which ensures adequate coverage near domain boundaries and improves approximation accuracy. By enumerating all such indices across scales, we obtain a multiscale wavelet system $\mathcal{D} = \{\psi_\lambda\}_{\lambda \in \Lambda}$, which is localized, isotropic, and directionally sensitive, capturing features that are poorly resolved by global spectral representations.

Dominant family identification. We seek coefficients $\mathbf{c} \in \mathbb{R}^{|\Lambda|}$ such that

$$\bar{u}(\mathbf{x}) \approx \sum_{\lambda \in \Lambda} c_\lambda \psi_\lambda(\mathbf{x}).$$

This is posed as a Tikhonov-regularized least-squares [18] problem:

$$\mathbf{c}^* = \arg\min_{\mathbf{c}} \|\mathcal{W}\mathbf{c} - \bar{u}\|^2_{L^2(\Omega)} + \lambda_{\text{reg}} \|\mathbf{c}\|^2,$$

whose normal equations are given by

$$(\mathcal{W}^T\mathcal{W} + \lambda_{\text{reg}} I)\mathbf{c} = \mathcal{W}^T \bar{u},$$

where $\mathcal{W}$ denotes the wavelet operator. The Gram operator $\mathcal{W}^T\mathcal{W} + \lambda_{reg} I$ is symmetric positive definite and is never formed explicitly. Instead, matrix-vector products are evaluated in a chunked, streaming manner, enabling the use of the conjugate gradient method [19, 20] for large-scale dictionaries without storing the full matrix. The right-hand side $\mathbf{b} = \mathcal{W}^T \bar{u}$ is computed via midpoint quadrature rule [21] on the uniform FNO evaluation grid, and convergence is monitored using the relative residual $\|\mathbf{r}_k\|/\|\mathbf{b}\|$, where r_k denotes the residual at the k-th iteration of the conjugate gradient method. Each conjugate-gradient iteration requires one dictionary analysis and one synthesis operation. Let N_{grid} denote the number of grid points, and n_{CG} the number of CG iterations. With chunked evaluation, the Stage II computational complexity is $\mathcal{O}(n_{\text{CG}}|\Lambda|N_{\text{grid}})$, with memory scaling as $\mathcal{O}(N_{\text{grid}}N_{\text{chunk}})$. In practice, n_{CG} remains small because Stage II is used only to obtain an approximate projection for subspace selection and coefficient initialization.

Wavelet subspace selection. We exploit sparsity by thresholding the spectral energy coefficients using the p-th empirical quantile, retaining only the dominant spectral basis and its corresponding coefficients:

$$\Lambda^* = \left\{\lambda \in \Lambda : |c_\lambda|^2 > \tau_p\big(\{|c_\lambda|^2\}_{\lambda \in \Lambda}\big)\right\},$$

where τ_p denotes the p-th quantile of the set $\{|c_\lambda|^2\}$. This yields a reduced, instance-adaptive subspace

$$\mathcal{V}^* = \text{span}\{\psi_\lambda\}_{\lambda \in \Lambda^*}, \quad |\Lambda^*| \ll |\Lambda|.$$

This subspace captures the dominant multiscale features of $\bar{u}$ at a fraction of the computational cost of the full system. Importantly, the index set Λ^* is instance-specific, depending on the input ε^*, and thus adapts the refinement basis to the intrinsic structure of the solution. The threshold τ_p is determined solely from the coarse prediction $\bar{u}$, without access to the reference solution, and controls only the sparsity of $\mathcal{V}^*$. We select the largest τ_p for which the reconstruction error

$$\frac{\|\sum_{\lambda \in \Lambda^*} c_\lambda \psi_\lambda - \bar{u}\|_{L^2(\Omega)}}{\|\bar{u}\|_{L^2(\Omega)}}$$

remains below a small tolerance, yielding a sparse admissible subspace.

Fourier dictionary. As an alternative formulation of Stage II, we construct a spectral subspace based on the discrete Fourier transform (DFT). This formulation captures global oscillatory structures and serves as a complementary representation to the localized multiscale basis. The coarse prediction $\bar{u}$ is decomposed into its spectral components, and the K dominant modes are retained to form the projection basis. This yields the F-AMOC, which shares Stages I and III identically with the W-AMOC and differs solely in replacing the spatially localized wavelet dictionary with a global spectral one. Concretely, let $\Omega = \prod_{i=1}^{n}[a_i, b_i]$ with side lengths $L_i = b_i - a_i$, and let $\bar{u}$ be evaluated on a uniform N^n-point grid. The n-dimensional DFT is defined as

$$\hat{U}_{\boldsymbol{\ell}} = \sum_{\mathbf{j}\in[N]^n} \bar{u}(\mathbf{x}^{(\mathbf{j})})\, e^{-2\pi i \boldsymbol{\ell}\cdot\mathbf{j}/N}, \qquad \boldsymbol{\ell} \in \mathbb{Z}^n,$$

with physical wavenumber vector, $\mathbf{k}_{\boldsymbol{\ell}} = \dfrac{2\pi\tilde{\boldsymbol{\ell}}}{\mathbf{L}}$, where $\tilde{\boldsymbol{\ell}}$ is the signed frequency index under the FFT convention and division is elementwise.

Sparse mode selection. The K modes with the largest spectral energy are retained:

$$\mathcal{K}^* = \underset{\mathcal{K}\subset\mathbb{Z}^n,\ |\mathcal{K}|=K}{\arg\max} \sum_{\boldsymbol{\ell}\in\mathcal{K}} |\hat{U}_{\boldsymbol{\ell}}|^2.$$

For each retained mode, the phase corrected for the domain origin is

$$\varphi_{\boldsymbol{\ell}} = -\mathbf{k}_{\boldsymbol{\ell}} \cdot \mathbf{a} + \arg(\hat{U}_{\boldsymbol{\ell}}), \qquad \mathbf{a} = (a_1, \ldots, a_n),$$

and the amplitude is $A_{\boldsymbol{\ell}} = \dfrac{|\hat{U}_{\boldsymbol{\ell}}|}{N^n}$. The sparse spectral subspace is $\mathcal{V}^*_{\mathcal{F}} = \operatorname{span}\{\cos(\mathbf{k}_{\boldsymbol{\ell}} \cdot \mathbf{x} + \varphi_{\boldsymbol{\ell}})\}_{\boldsymbol{\ell}\in\mathcal{K}^*}$.

2.5 Stage III: Physics-constrained optimization

Given the sparse index set Λ^* and initialized coefficients $\{c_\lambda\}_{\lambda\in\Lambda^*}$, we construct a parametric model within $\mathcal{V}^*$ and enforce physical consistency by minimizing the PDE residual.

Adaptive wavelet model. Let $M = |\Lambda^*|$ denote the number of selected basis functions. For $k = 1, 2, \ldots, M$, we parameterize the solution as

$$u_\theta(\mathbf{x}) = \sum_{k=1}^{M} c_k\, \psi_k(\mathbf{x};\, \mathbf{w}_k, \mathbf{b}_k) + b_0,$$

where $\mathbf{x} = (x_1, \ldots, x_d) \in \Omega \subset \mathbb{R}^d$, and for each wavelet $\mathbf{w}_k = (w_k^1, \ldots, w_k^d)$, $\mathbf{b}_k = (b_k^1, \ldots, b_k^d)$ and corresponding coefficients c_k are trainable parameters initialized from Stage II. Define

$$z_k^i = w_k^i\, x_i + b_k^i, \quad i = 1, \ldots, d.$$

Then, the d-dimensional wavelet functions are given by

$$\psi_k(\mathbf{x}) = \left(\prod_{i=1}^{d} \sqrt{w_k^i}\right) \prod_{i=1}^{d} \left(-z_k^i\, e^{-(z_k^i)^2/2}\right).$$

Let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}^d$ be a multi-index. The $\boldsymbol{\alpha}$-th order derivative of ψ_k is given by

$$\partial^{\boldsymbol{\alpha}}\psi_k(\mathbf{x}) = \left(\prod_{i=1}^{d} \sqrt{w_k^i}\right) \prod_{i=1}^{d} \left[(w_k^i)^{\alpha_i}\, (-1)^{\alpha_i+1} H_{\alpha_i+1}(z_k^i)\, e^{-(z_k^i)^2/2}\right],$$

where $z_k^i = w_k^i x_i + b_k^i$, and $H_n(\cdot)$ denotes the Hermite polynomial of order n. Since all basis functions admit closed-form derivatives, no automatic differentiation through computational graphs is required, significantly reducing computational overhead.

Adaptive Fourier model. Given the sparse index set Λ^* with $M = |\Lambda^*|$, we construct a parametric model using adaptive Fourier features. The solution is represented as

$$u_\theta(\mathbf{x}) = \sum_{k=1}^{M} c_k \, \cos\big(\mathbf{w}_k \cdot \mathbf{x} + \phi_k\big) + b_0,$$

where $\mathbf{x} = (x_1, \ldots, x_d) \in \Omega \subset \mathbb{R}^d$. Each mode is parameterized by a frequency vector $\mathbf{w}_k = (w_k^1, \ldots, w_k^d) \in \mathbb{R}^d$, a phase shift $\phi_k \in \mathbb{R}$ which implicitly accounts for sine terms through phase shifts, and coefficient $c_k \in \mathbb{R}$, all initialized from Stage II. The bias term $b_0 \in \mathbb{R}$ captures the mean component. The $\boldsymbol{\alpha}$-th order partial derivative of u_θ is given by

$$\partial^{\boldsymbol{\alpha}} u_\theta(\mathbf{x}) = \sum_{k=1}^{M} c_k \left(\prod_{i=1}^{d} (w_k^i)^{\alpha_i} \right) \cos\Big(\mathbf{w}_k \cdot \mathbf{x} + \phi_k + \frac{\pi}{2}\boldsymbol{\alpha}_{\boldsymbol{s}}\Big), \text{ where } \boldsymbol{\alpha}_s = \sum_{i=1}^{d} \alpha_i.$$

Remarks. Similar to the wavelet construction, the Fourier basis enables exact evaluation of arbitrary-order derivatives without relying on automatic differentiation. This results in improved numerical stability and reduced computational complexity, particularly for higher-order PDE constraints.

Physics-informed training. At collocation points $\{\mathbf{x}_i\}_{i=1}^{N_c}$, the PDE residual is defined as

$$\mathcal{R}(\mathbf{x};\, \theta) = \mathcal{N}[u_\theta;\, \varepsilon](\mathbf{x}) - f(\mathbf{x}).$$

The training objective enforces physics consistency along with boundary and initial conditions:

$$\mathcal{L}(\theta) = \alpha \underbrace{\frac{1}{N_c} \sum_{i=1}^{N_c} |\mathcal{R}(\mathbf{x}_i;\, \theta)|^2}_{\mathcal{L}_{\text{PDE}}} + \beta \underbrace{\frac{1}{N_B} \sum_{k=1}^{N_B} |\mathcal{B}[u_\theta](\mathbf{x}_k) - g(\mathbf{x}_k)|^2}_{\mathcal{L}_{\text{BC}}} + \gamma \underbrace{\frac{1}{N_0} \sum_{j=1}^{N_0} |u_\theta(\mathbf{x}_j) - u_0(\mathbf{x}_j)|^2}_{\mathcal{L}_{\text{IC}}},$$

where α, β, γ is a weighting parameter and N_c, N_B, N_0 denotes the number of interior collocation points, boundary points, and initial condition points, respectively. The optimal parameters are obtained via

$$\theta^* = \arg\min_{\theta} \mathcal{L}(\theta),$$

typically using a second-order optimizer such as L-BFGS. We use L-BFGS because Stage III is initialised with the coefficients obtained from Stage II rather than a random initialisation. The resulting optimization is performed over the relatively low-dimensional adaptive subspace. The reduced dimensionality of the adaptive subspace ($M \ll |\Lambda|$) enables efficient optimization with rapid convergence.

3 Experiments

All experiments are conducted on an NVIDIA RTX PRO 4500 Blackwell GPU with CUDA 12.3. To ensure reproducibility, we fix the random seeds across all runs. The FNO baseline is trained on 400 labeled (ε, u) pairs sampled uniformly over the domain Ω. The data is prepared at resolution 128×128 and $128 \times 128 \times 128$, and subsequently downsampled by a factor of two for 2D and 3D problems, respectively. The optimization is performed using the Adam optimizer with a learning rate of 10^{-3} and weight decay of 10^{-5}. A StepLR scheduler with step size 100 and decay factor $\gamma = 0.75$ is applied over 500 epochs with the batch size 10. The model architecture consists of four Fourier layers with width 32 and 20 spectral modes, unless stated otherwise. The Stage III optimization of W-AMOC and F-AMOC is carried out using L-BFGS. Both bases admit closed-form derivatives via Hermite polynomials for W-AMOC and cosine phase shifts for F-AMOC, allowing the analytical evaluation of derivatives involved in PDE residuals, which reduces computational cost by avoiding the need for automatic differentiation. We retain the top 10% dominant wavelet bases and 200 Fourier modes throughout the experiments, unless otherwise specified. A detailed hyperparameter table is provided in Appendix C. Performance is evaluated using the relative $\mathcal{L}_2$-error metric, on 10 in-distribution (ID) and 10 out-of-distribution (OOD) samples. We report the mean relative $\mathcal{L}_2$-error and standard deviation across these samples in Table 1, 2. The source code and data used in this study can be found at this https://github.com/subhampatel2899/AMOC.

3.1 Darcy flow

As a canonical benchmark in PDE modeling, flow through heterogeneous porous media is widely studied. In this work, we consider the two-dimensional variable-coefficient Darcy flow equation

$$\begin{cases} -\nabla \cdot \big(k(x,y)\,\nabla p(x,y)\big) = f(x,y), & (x,y) \in \Omega \equiv [0,1]^2, \\ p(x,y) = 0, & (x,y) \in \partial\Omega, \end{cases}$$

where the permeability field, k, is a superposition of five Gaussian bumps,

$$k(x,y) = 1 + \sum_{j=1}^{5} A_j \exp\left(-\frac{(x-x_0^j)^2 + (y-y_0^j)^2}{2\sigma_j^2}\right),$$

and the forcing, f is a superposition of sinusoidal modes,

$$f(x,y) = \sum_{m=1}^{5} B_m \, \sin(n_x^m \pi x) \, \sin(n_y^m \pi y).$$

For the ID setting, parameters are drawn as $A_j \sim \mathcal{U}(2,8)$, $\sigma_j \sim \mathcal{U}(0.10, 0.20)$, $B_m \sim \mathcal{U}(-5,5)$, $x_0^j, y_0^j \sim \mathcal{U}(0,1)$ and $n_x^m, n_y^m \sim \mathcal{U}\{1,\dots,7\}$, while the OOD regime employs sharper permeability fields with $\sigma_j \sim \mathcal{U}(0.05, 0.10)$, $A_j \sim \mathcal{U}(8,12)$, and stronger forcing $B_m \sim \mathcal{U}(-10,10)$, yielding more localized pressure gradients unseen during training. The ground-truth pressure fields are computed via a Gauss-Seidel finite-difference solver with up to 5×10^5 iterations and convergence tolerance 10^{-6} [6]. The solution operator $\mathcal{G}_\theta : (k, f) \mapsto p$ maps the joint permeability-forcing input to the full pressure field.

3.2 Heat equation

As a prototypical model for time-dependent diffusion processes, we consider the heat equation in the following parabolic PDE form:

$$\begin{cases} u_t(x,t) - \varepsilon\, u_{xx}(x,t) = f(x,t), & (x,t) \in \Omega \equiv [-1,1] \times [0,1], \\ u(-1,t) = u(1,t) = 0, & t \in [0,1], \\ u(x,0) = (1-x^2)\exp(1/(1+\varepsilon)), & x \in [-1,1], \end{cases}$$

where f is chosen such that the exact solution is given by

$$u(x,t) = (1-x^2)\exp\left(\frac{1}{(2t-1)^2 + \varepsilon}\right).$$

For small ε, the solution develops a sharp exponential peak concentrated near $t = 0.5$, making accurate approximation particularly challenging for data-driven methods trained on coarse grids. The solution operator $\mathcal{G}_\theta : \varepsilon \mapsto u$ maps the constant field ε to the corresponding space-time solution field. We use two test sets, an ID set with $\varepsilon \sim \mathcal{B}(0.6, 1.4)$ scaled to $[0.15, 0.5]$, and an OOD set scaled to $[0.10, 0.15]$, producing sharper temporal gradients unseen during training.

3.3 Poisson equation

The Poisson equation is a fundamental elliptic PDE arising in steady-state diffusion and potential theory. In this study, we consider the d-dimensional Poisson equation

$$\begin{cases} -\Delta u(\mathbf{x}) = f(\mathbf{x}), & \mathbf{x} \in \Omega \equiv [0,1]^d, \\ u(\mathbf{x}) = g(\mathbf{x}), & \mathbf{x} \in \partial\Omega, \end{cases}$$

In 2D setting, we consider solutions of the form

$$u(x,y) = 1 + (1000 + y^2)\exp\left(-\frac{(x-0.5)^2}{2\varepsilon^2}\right), \tag{2}$$

with g obtained by evaluating (2) on $\partial\Omega$ and f derived analytically. For small ε, the solution develops a sharp boundary layer near $x = 0.5$, making accurate approximation particularly challenging. The

solution operator $\mathcal{G}_\theta : \varepsilon \mapsto u$ maps the scalar parameter to the full spatial field. We use two test sets, first an ID set with $\varepsilon \sim \mathcal{B}(0.6, 1.4)$ scaled to $[0.03, 0.06]$, and an OOD set with $\varepsilon \sim \mathcal{U}[0.015, 0.03]$, inducing sharper boundary layers unseen during training.

In 3D settings, solutions are constructed as a superposition of isotropic Gaussians

$$u(x, y, z) = \sum_{k=1}^{20} A_k \exp\big(-\alpha_k\, r_k^2\big)\,, \quad r_k^2 = (x - x_0^k)^2 + (y - y_0^k)^2 + (z - z_0^k)^2, \tag{3}$$

with g obtained by evaluating (3) on $\partial\Omega$ and $f = \sum_k A_k \exp(-\alpha_k r_k^2)(6\alpha_k - 4\alpha_k^2 r_k^2)$ derived analytically. For large α_k, the Gaussians become sharply peaked, producing strongly localized features challenging for global spectral representations. The solution operator $\mathcal{G}_\theta : f \mapsto u$ maps instance parameters to the full solution field. We use two test sets, an ID set with $A_k \sim \mathcal{U}(-1, 1)$, $(x_0^k, y_0^k, z_0^k) \sim \mathcal{U}(0.1, 0.9)^3$, $\alpha_k \sim \mathcal{U}(60, 80)$ and an OOD set with $\alpha_k \sim \mathcal{U}(80, 100)$, producing sharper concentrated Gaussians unseen during training.

3.4 Helmholtz equation

The Helmholtz equation models wave propagation and is a standard PDE exhibiting high-frequency oscillatory behaviour. We focus on the three-dimensional setting

$$\begin{cases} \Delta u(x, y, z) + k^2 u(x, y, z) = f(x, y, z), \quad (x, y, z) \in \Omega \equiv [-1, 1]^3, \\ u(x, y, z) = 0, \quad (x, y, z) \in \partial\Omega, \end{cases}$$

where the forcing term $f = \big(k^2 - (2\pi)^2(b_1^2 + b_2^2 + b_3^2)\big)u(x, y, z)$ derived analytically from the Helmholtz operator using the analytical solution:

$$u(x, y, z) = \sin(2\pi b_1 x)\, \sin(2\pi b_2 y)\, \sin(2\pi b_3 z).$$

For large frequency modes b_i, the solution exhibits rapid spatial oscillations, leading to high-frequency interference patterns that make this problem a challenging test case for neural operators. The solution operator $\mathcal{G}_\theta : (f, k) \mapsto u$ maps the forcing field and wave number to the full solution field. We use two test sets, an ID set with $b_i \sim \mathcal{U}\{1, \ldots, 5\}$ and $k \sim \mathcal{U}(0.5, 5.5)$, and an OOD set with $b_i \sim \mathcal{U}\{4, 5, 6\}$ and $k \sim \mathcal{U}(2, 7)$, producing higher-frequency, more oscillatory solutions unseen during training.

Table 1: Relative $\mathcal{L}_2$-error (mean $\pm$ standard deviation) for in-distribution settings across multiple PDE benchmarks

	FNO	PINO	PINO correction	F-AMOC	W-AMOC
Darcy flow	1.32e-01 ± 7.81e-04	1.27e-01 ± 2.55e-02	4.78e-02 ± 1.89e-02	8.66e-01 ± 3.65e-01	4.16e-02 ± 6.90e-02
Heat eqn	7.00e-03 ± 3.68e-03	3.09e-03 ± 1.95e-03	3.29e-03 ± 8.51e-04	2.39e-03 ± 2.37e-03	5.85e-05 ± 2.00e-05
Poisson eqn	4.30e-03 ± 7.03e-04	4.44e-03 ± 1.06e-03	3.66e-03 ± 9.15e-04	4.88e-04 ± 2.84e-04	9.03e-04 ± 1.22e-03
3D Poisson eqn	7.66e-02 ± 1.17e-02	6.43e-02 ± 2.07e-02	1.74e-02 ± 5.02e-04	3.43e-01 ± 5.09e-02	4.17e-03 ± 3.76e-03
3D Helmholtz eqn	7.27e-03 ± 2.46e-03	7.55e-03 ± 2.23e-03	3.37e-02 ± 1.15e-02	1.55e-05 ± 1.17e-05	6.71e-02 ± 2.00e-02

4 Results

The relative $\mathcal{L}_2$-error for the baseline FNO, PINO, PINO correction and the corrective models, F-AMOC and W-AMOC, are reported in Tables 1, 2. PINO and PINO correction provide well-established physics-informed baselines, while F-AMOC and W-AMOC are designed as corrective approaches to further improve the accuracy of the underlying operator predictions. For problems characterized by localized features, such as Heat, Poisson, and Darcy, the localized basis employed

Table 2: Relative $\mathcal{L}_2$-error (mean $\pm$ standard deviation) for out-of-distribution settings across multiple PDE benchmarks

	FNO	PINO	PINO correction	F-AMOC	W-AMOC
Darcy flow	2.46e-01 $\pm$ 6.49e-03	2.12e-01 $\pm$ 4.02e-02	5.00e-02 $\pm$ 6.90e-02	4.93e-01 $\pm$ 2.26e-01	2.84e-02 $\pm$ 2.06e-02
Heat eqn	3.45e-01 $\pm$ 1.81e-01	3.35e-01 $\pm$ 1.91e-01	8.37e-03 $\pm$ 2.82e-03	6.04e-03 $\pm$ 5.48e-03	3.79e-05 $\pm$ 1.07e-05
Poisson eqn	1.53e-01 $\pm$ 1.39e-01	1.03e-01 $\pm$ 9.90e-02	1.14e-02 $\pm$ 6.19e-03	3.28e-02 $\pm$ 4.24e-02	9.27e-04 $\pm$ 2.04e-04
3D Poisson eqn	6.70e-02 $\pm$ 8.64e-03	5.72e-02 $\pm$ 1.33e-02	1.97e-02 $\pm$ 1.50e-03	4.10e-01 $\pm$ 2.53e-02	6.71e-03 $\pm$ 1.13e-02
3D Helmholtz eqn	1.09e-01 $\pm$ 3.77e-02	7.58e-02 $\pm$ 7.36e-02	6.13e-02 $\pm$ 1.18e-02	3.11e-05 $\pm$ 6.22e-05	6.98e-02 $\pm$ 2.80e-02

in W-AMOC leads to superior performance compared to F-AMOC. In contrast, for the Helmholtz problem, which exhibits a global oscillatory structure, F-AMOC outperforms W-AMOC.

A qualitative comparison for the Darcy problem (3.1) is presented in Fig. 2. The pointwise error distribution shows that W-AMOC attains lower error across the domain. This observation is further reinforced by the spatial snapshot plots. For other examples, solutions and their error plots are provided in Appendix A. Further, Fig. 3 shows a bar plot of model accuracy with the number of selected basis functions for 2D problems. The results indicate that both methods consistently improve upon the baseline accuracy as the number of basis functions increases. With a sufficient number of basis functions, W-AMOC achieves higher accuracy than F-AMOC. We also report statistical measures of the AMOC model in Table 3 and Fig. 9 in Appendix A. Table 3 presents the mean relative $\mathcal{L}_2$-error and standard deviation across 10 (5) random runs for 2D (3D) problems, where

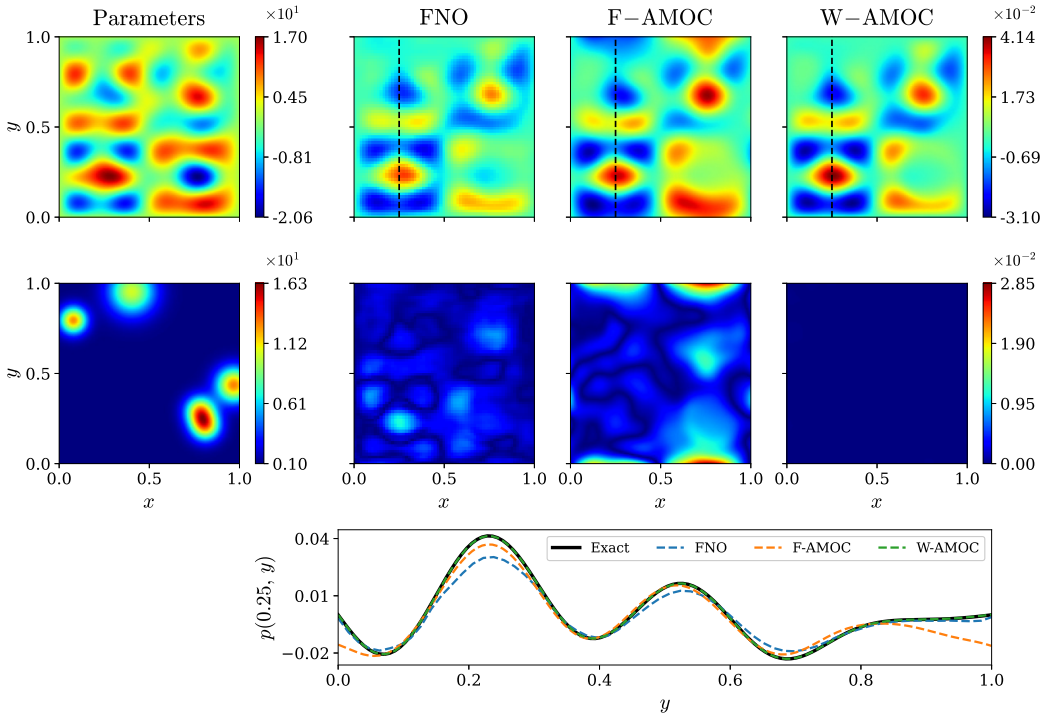

Figure 2: Solution of Darcy flow (3.1) for various models and their respective point-wise errors.
Left panel: Input parameter fields - forcing field (top) and permeability field (bottom).
Right panel (top): Predictions via FNO, F-AMOC, and W-AMOC. Middle: Corresponding point-wise error. Bottom: Spatial snapshot plot at $x = 0.25$, along with the reference solution.

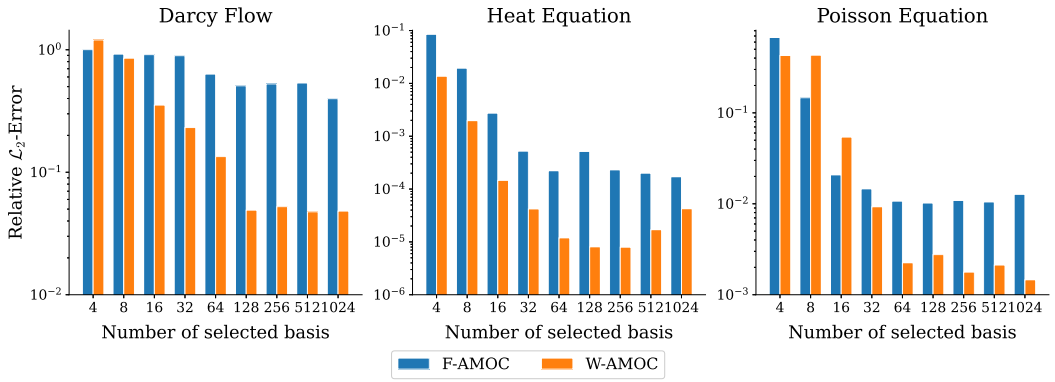

Figure 3: Bar plot of relative $\mathcal{L}_2$-error versus the number of selected basis functions for W-AMOC (orange) and F-AMOC (blue). From left to right: Darcy flow (3.1), Heat equation (3.2), and 2D Poisson equation (3.3).

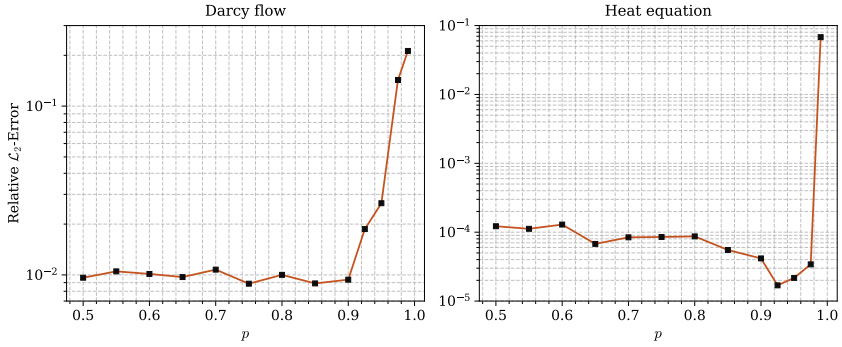

Figure 4: Relative $\mathcal{L}_2$-error of W-AMOC versus the Stage II quantile p for Darcy flow (left) and the heat equation (right).

lower standard deviation indicates model's stability and low variance. We have also presented the sensitivity analysis of the W-AMOC with respect to the quantile level p in Fig. 4 which shows that accuracy is nearly unchanged over a broad range ($p \approx 0.5$ - 0.9) and degrades only as $p \to 1$, when the subspace becomes overly sparse. For the 3D problems, the higher values ($p = 0.98$-0.99, Table 7) are set by the memory budget of the 3D wavelet dictionary. In addition, we evaluate computational efficiency by reporting total training time and memory usage during the correction stage in Table 4.

5 Limitations and Conclusion

Our proposed method introduces an adaptive spectral correction framework that refines a coarse neural operator prediction into a high-fidelity, physics-consistent solution by performing instance-specific spectral physics-informed optimization. Experimental results demonstrate significant accuracy improvement across diverse PDE benchmarks, particularly in OOD regimes. The comparison between wavelet and Fourier bases highlights the importance of choosing representations aligned with problem characteristics. While wavelet representations offer advantages for capturing localized features, their memory requirements grow rapidly when extending from 2D to 3D. In contrast, Fourier bases are more memory-efficient but struggle with localized structures. This trade-off suggests that hybrid approaches combining Fourier and wavelet representations could provide a more balanced solution. Overall, AMOC provides a scalable and accurate paradigm for enhancing neural operator performance in complex multiscale systems.

Acknowledgement

Subham Patel gratefully acknowledges the Axis Bank Centre for Mathematics and Computing, IISc, Bengaluru, and the Ministry of Education, Government of India, for fellowship and contingency support, and the Pratiksha Trust Initiatives, IISc, Bengaluru, for the travel fellowship to attend and present this work at NeurIPS 2026.

Himanshu Pandey gratefully acknowledges Google from Kotak IISc AI-ML Centre for providing financial support to attend and present this work at the NeurIPS conference. The author also acknowledges the Ministry of Education, Government of India, for fellowship support.

We are also grateful to the anonymous reviewers for their valuable comments and suggestions, which have helped us improve the quality of this work.

References

[1] Lawrence C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2010. doi: 10.1090/gsm/019.

[2] John F. Wendt. *Computational fluid dynamics*. A von Karman Institute Book. Springer-Verlag, Berlin, third edition, 2009. doi: 10.1007/978-3-540-85056-4.

[3] Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: learning maps between function spaces with applications to PDEs. *J. Mach. Learn. Res.*, 24(1), January 2023. ISSN 1532-4435.

[4] Tanya Marwah, Ashwini Pokle, J Zico Kolter, Zachary Lipton, Jianfeng Lu, and Andrej Risteski. Deep equilibrium based neural operators for steady-state PDEs. *Advances in Neural Information Processing Systems*, 36:15716–15737, 2023.

[5] 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.

[6] Zongyi Li, Nikola Borislavov 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://openreview.net/forum?id=c8P9NQVtmnO.

[7] Gaurav Gupta, Xiongye Xiao, and Paul Bogdan. Multiwavelet-based operator learning for differential equations. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 24048–24062. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/c9e5c2b59d98488fe1070e744041ea0e-Paper.pdf.

[8] Alasdair Tran, Alexander Mathews, Lexing Xie, and Cheng Soon Ong. Factorized Fourier neural operators, 2023. URL https://arxiv.org/abs/2111.13802.

[9] Shuhao Cao. Choose a transformer: Fourier or Galerkin. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 24924–24940. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/d0921d442ee91b896ad95059d13df618-Paper.pdf.

[10] Salvatore Cuomo, Vincenzo Schiano Di Cola, Fabio Giampaolo, Gianluigi Rozza, Maziar Raissi, and Francesco Piccialli. Scientific machine learning through physics–informed neural networks: Where we are and what's next. *Journal of Scientific Computing*, 92(3):88, 2022.

[11] 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 *International conference on machine learning*, pages 5301–5310. PMLR, 2019.

[12] Zhi-Qin John Xu, Yaoyu Zhang, Tao Luo, Yanyang Xiao, and Zheng Ma. Frequency principle: Fourier analysis sheds light on deep neural networks. *Communications in Computational Physics*, 28(5):1746–1767, 2020. ISSN 1991-7120. doi: 10.4208/cicp.oa-2020-0085. URL http://dx.doi.org/10.4208/cicp.OA-2020-0085.

[13] Armand Kassaï Koupaï, Jorge Mifsut Benet, Yuan Yin, Jean-Noël Vittaut, and Patrick Gallinari. Boosting generalization in parametric PDE neural solvers through adaptive conditioning. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang, editors, *Advances in Neural Information Processing Systems*, volume 37, pages 70659–70692. Curran Associates, Inc., 2024. doi: 10.52202/079017-2258. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/82844e428d9163a9f94830dc03af4f9c-Paper-Conference.pdf.

[14] Yong Wang, Yanzhong Yao, Jiawei Guo, and Zhiming Gao. A practical PINN framework for multi-scale problems with multi-magnitude loss terms. *Journal of Computational Physics*, 510:113112, 2024. ISSN 0021-9991. doi: https://doi.org/10.1016/j.jcp.2024.113112. URL https://www.sciencedirect.com/science/article/pii/S0021999124003619.

[15] Himanshu Pandey, Anshima Singh, and Ratikanta Behera. An efficient wavelet-based physics-informed neural network for multiscale problems. *Neural Networks*, page 108860, 2026.

[16] Himanshu Pandey and Ratikanta Behera. An adaptive wavelet-based PINN for problems with localized high-magnitude source, 2026. URL https://arxiv.org/abs/2604.28180.

[17] Zongyi Li, Hongkai Zheng, Nikola Kovachki, David Jin, Haoxuan Chen, Burigede Liu, Kamyar Azizzadenesheli, and Anima Anandkumar. Physics-informed neural operator for learning partial differential equations. *ACM/IMS Journal of Data Science*, 1(3):1–27, 2024.

[18] Gene H Golub and Charles F Van Loan. *Matrix computations*. JHU press, 2013.

[19] Magnus R Hestenes, Eduard Stiefel, et al. Methods of conjugate gradients for solving linear systems. *Journal of research of the National Bureau of Standards*, 49(6):409–436, 1952.

[20] Jonathan Richard Shewchuk et al. An introduction to the conjugate gradient method without the agonizing pain, 1994.

[21] Lloyd N Trefethen. *Spectral methods in MATLAB*. SIAM, 2000.

A Additional results and discussions

In this section, we provide detailed visualizations and supplementary quantitative results for all benchmark problems. For each equation, we present the predicted solution fields obtained by FNO, F-AMOC, and W-AMOC alongside the corresponding pointwise absolute errors, separately for in-distribution and out-of-distribution test instances.

A.1 Heat equation

In the OOD regime, we consider a constant field $\varepsilon = 0.1$, the solution develops a sharply localized exponential peak near $t = 0.5$, with rapid variation confined to a narrow temporal region and smooth behavior elsewhere (see Fig. 5). FNO captures the overall trend but fails to resolve this localized structure due to its reliance on global Fourier modes, leading to oversmoothing and large pointwise errors around the peak.

F-AMOC operates in an adaptive spectral subspace, improving the approximation by selecting dominant modes. However, the non-local nature of the basis limits its ability to represent sharp localization, and residual error persists near the peak. In contrast, W-AMOC performs correction in a spatially localized wavelet subspace constructed from the FNO prediction. This enables selective correction in regions of high variation through adaptive use of the dominant basis, allowing accurate capture of the peak's location and sharpness. As a result, W-AMOC substantially reduces the error and achieves the lowest pointwise error. This improvement is also reflected in Table 3 and illustrated in Fig. 9, where W-AMOC attains lower error and reduced variance compared to F-AMOC for a representative OOD instance.

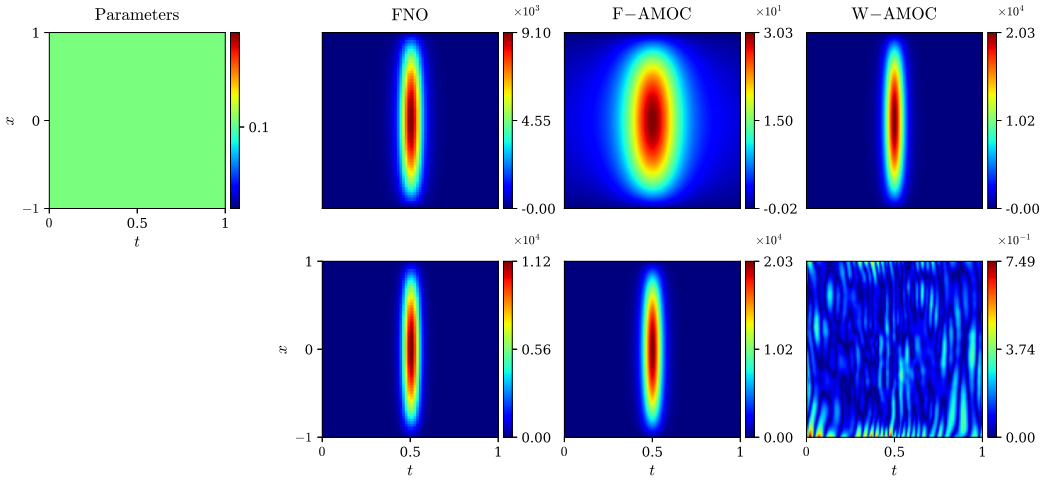

Figure 5: Left panel: Input constant parameter field ($\varepsilon = 0.1$). Right panel: Top: Predictions of Heat equation (3.2) via FNO, F-AMOC, and W-AMOC. Bottom: Corresponding point-wise error.

Table 3: Mean and standard deviation of relative $\mathcal{L}_2$-error for Stage III across PDE benchmarks on a single representative OOD instance, evaluated over 10 (5) independent runs for 2D (3D) problems.

Equation	F-AMOC	W-AMOC
Darcy flow	(5.30 ± 0.21)e-1	(4.97 ± 0.63)e-3
Heat eqn	(1.91 ± 0.04)e-3	(1.50 ± 0.30)e-5
2D Poisson eqn	(1.83 ± 0.92)e-2	(8.63 ± 2.91)e-4
3D Poisson eqn	(3.95 ± 0.10)e-1	(5.09 ± 0.27)e-3
3D Helmholtz eqn	(1.50 ± 0.90)e-5	(7.11 ± 0.07)e-2

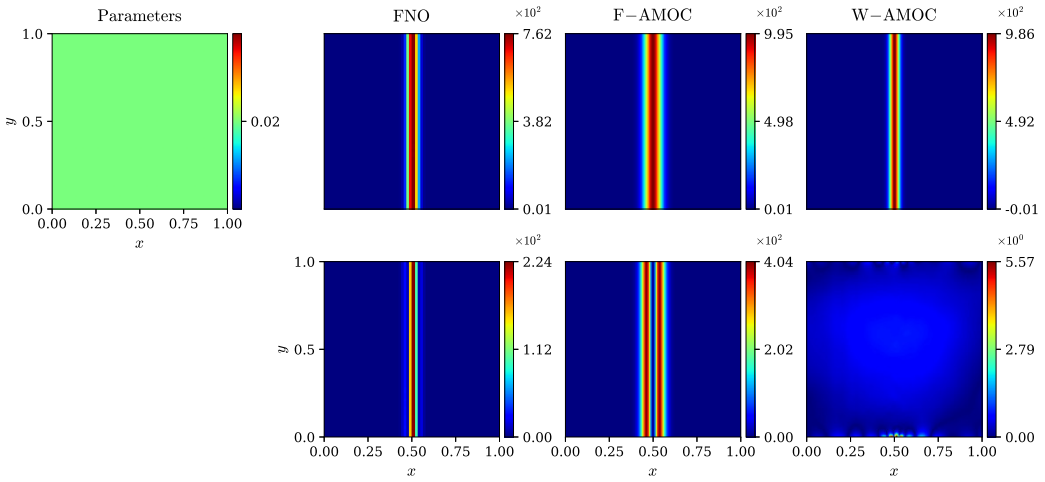

Figure 6: Left panel: Input constant parameter field ($\varepsilon = 0.02$). Right panel: Top: Predictions of Poisson equation (2) via FNO, F-AMOC, and W-AMOC. Bottom: Corresponding point-wise error.

A.2 Poisson equation

In the 2D setting, Fig. 6 presents results for the Poisson equation at $\varepsilon = 0.02$, corresponding to an OOD regime. The solution exhibits a sharp boundary layer localized around $x = 0.5$, characterized by rapid variation across a narrow spatial region and near-constant behavior elsewhere. FNO struggles to resolve this feature due to its reliance on global Fourier modes, which are inefficient for representing localized transitions. This results in an overly smoothed solution and a spread of pointwise error across the layer region.

F-AMOC mitigates this issue by correcting the FNO prediction in an adaptive spectral subspace, yielding a sharper approximation and reduced overall error. However, the use of global basis functions still limits its ability to fully capture the steep gradients, leaving noticeable error along the layer. In contrast, W-AMOC employs spatially localized Gaussian derivative wavelets, which better capture localized patterns and fine-scale variations, leading to more accurate predictions. Quantitatively, Table 3 shows that W-AMOC achieves lower error than F-AMOC, while Fig. 9 illustrates the variability across multiple runs for a representative OOD instance, showing low variance. In the 3D setting, the solution is composed of 20 isotropic Gaussians. Under the OOD

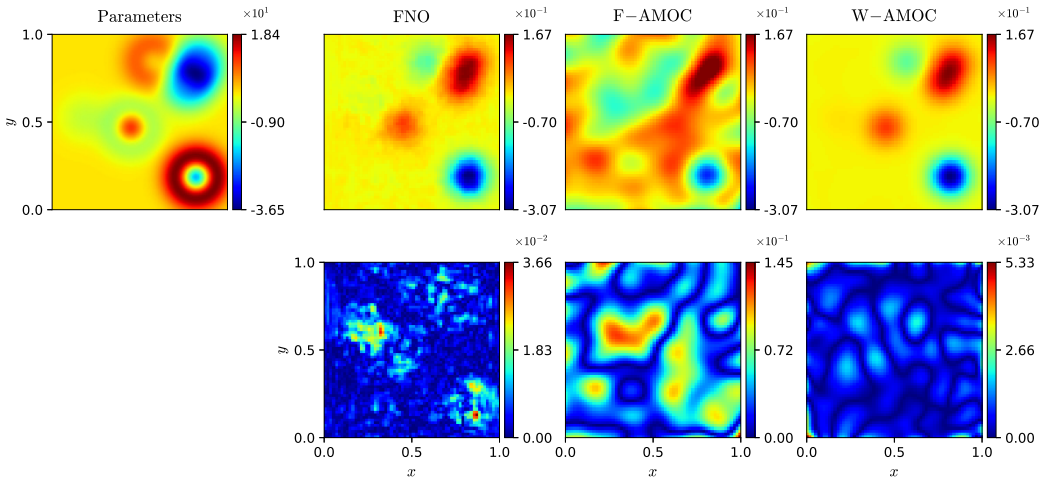

Figure 7: Left panel: Input forcing field. Right panel: Top: Predictions of 3D Poisson equation (3) via FNO, F-AMOC, and W-AMOC. Bottom: Corresponding point-wise error.

regime ($\alpha_k \sim \mathcal{U}(80, 100)$), these components become highly localized and exhibit strong directional variation. FNO captures the coarse structure but smooths out fine-scale details and introduces noticeable artifacts, as visualized in Fig. 7.

F-AMOC, despite performing correction in an adaptive spectral subspace, is unable to effectively rectify these errors and produces distorted structures, particularly in regions with strong localization. On the other hand, W-AMOC performs correction in a localized wavelet subspace and selects a sparse set of dominant wavelet basis based on energy. This allows it to better capture both the position and isotropic shape of the Gaussian components, resulting in a more accurate solution with significantly reduced error. For this problem, due to the high computational cost associated with the 3D wavelet dictionary, the correction is performed with fixed basis functions, optimizing only the spectral coefficients within the selected subspace. As shown in Fig. 9, W-AMOC consistently outperforms F-AMOC with lower variance, demonstrating its effectiveness for highly localized solutions for one specific instance. This observation is further supported by Table 3 and Fig. 9, which are computed on a representative OOD instance, where W-AMOC outperforms F-AMOC for this highly localized problem.

A.3 3D Helmholtz equation

The 3D Helmholtz equation is considered in an OOD regime with $b_i \in \{4, 5, 6\}$ and $k \sim \mathcal{U}(2, 7)$. The solution exhibits globally oscillatory behavior without spatial localization. FNO captures the overall oscillatory structure but leaves noticeable residual errors due to limited spectral resolution. F-AMOC corrects the FNO prediction by optimizing over a sparse set of relevant Fourier frequencies, leading to a substantial reduction in error, over three orders of magnitude.

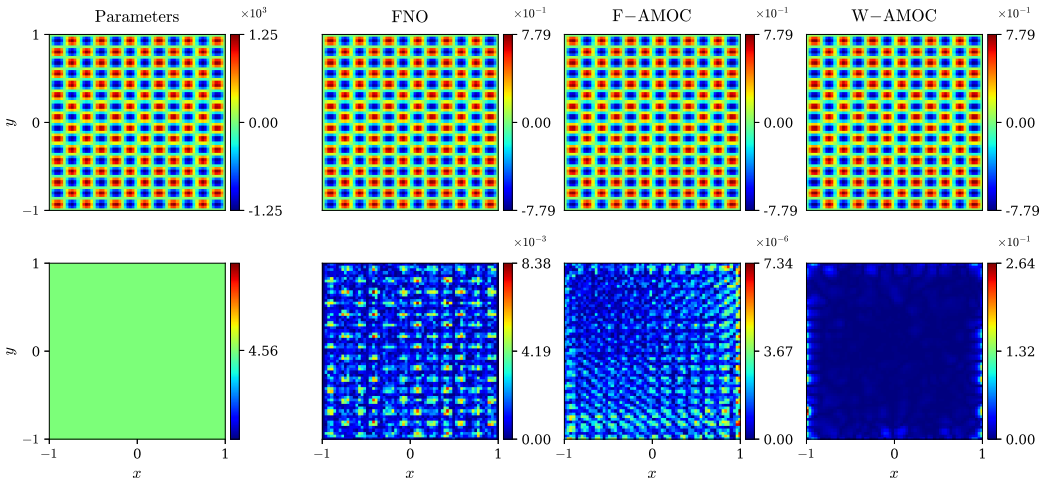

Figure 8: Left panel: Input fields (k, f) (representative z-slice, OOD). Right panel - Top: Predictions of the 3D Helmholtz equation (3.4) via FNO, F-AMOC, and W-AMOC. Bottom: Corresponding pointwise absolute errors. F-AMOC outperforms W-AMOC on this globally oscillatory problem.

In contrast, W-AMOC relies on localized wavelet bases, which are less efficient for representing globally oscillatory patterns, resulting in comparatively higher residual errors. These observations, illustrated in Fig. 8, highlight the complementary strengths of the two approaches i.e. F-AMOC is well-suited for globally oscillatory solutions, whereas W-AMOC is more effective for spatially localized features. This trend is also visible in Table 3 and Fig. 9, evaluated on a representative OOD instance, where F-AMOC achieves significantly lower error compared to W-AMOC.

B Further analysis and discussion

Accuracy-cost trade-off. As a post-processing framework, AMOC systematically corrects neural operator predictions. Although Stages II and III incur additional per-instance computation, this trade-off is highly effective. As illustrated for a representative out-of-distribution instance of the 2D

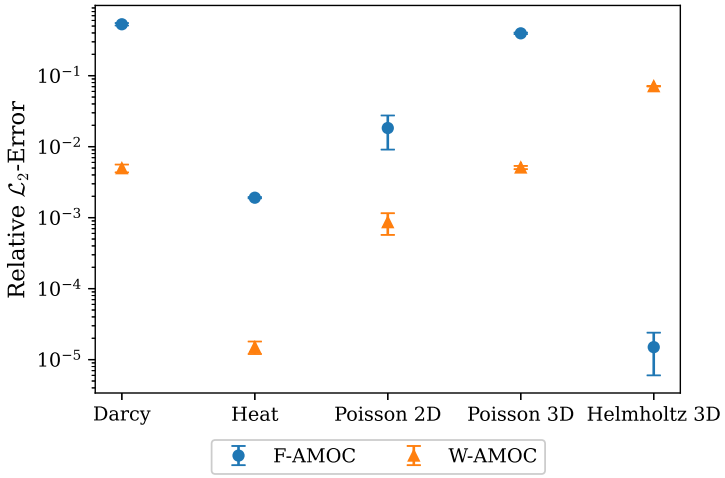

Figure 9: Relative $\mathcal{L}_2$-error for OOD single-instance correction using F-AMOC and W-AMOC across different PDE benchmarks. Markers indicate mean error, while error bars represent standard deviation over independent trials (10 runs for 2D problems and 5 runs for 3D problems).

Poisson equation in Figure 10, W-AMOC leverages this extra optimization time to achieve substantial accuracy gains over PINO correction. The computational cost is higher for the 3D problems in the current implementation, particularly for W-AMOC. Efficient parallelization and sparse evaluation of the multiscale dictionary are promising avenues to reduce this overhead.

Choice of spectral dictionary. Our results demonstrate that F-AMOC and W-AMOC offer complementary representational strengths tailored to distinct PDE characteristics. F-AMOC is well-suited to globally oscillatory solutions such as 3D Helmholtz, where the solution is sparse in the Fourier basis. W-AMOC is better suited to localized and multiscale features because its locally supported basis provide a sparse representation. The heat and Poisson solutions are not exactly represented by either finite dictionary, and the observed improvements therefore reflect the correction within an approximate subspace. The Darcy problem illustrates a limitation of energy-based Fourier selection. Its localized features are distributed over many low-energy Fourier modes. Energy thresholding can therefore remove modes that are collectively important for the localized structure. Consequently, Stage III is restricted to a subspace that cannot fully resolve these features, which can cause F-AMOC to degrade an already accurate FNO prediction. W-AMOC circumvents this issue, as its wavelet bases can sparsely and efficiently encode these localized structures. This contrasting behavior motivates the exploration of residual-aware or hybrid basis selection strategies in future work.

Conditioning and Stage III optimization. The wavelet dictionary is overcomplete, and correlations between basis functions can make its Gram operator ill-conditioned. The Tikhonov term makes $\mathcal{W}^\top\mathcal{W} + \lambda_{\text{reg}} I$ symmetric positive definite and improves its conditioning. Stage II only needs an approximate projection to select the basis and initialize the coefficients, so a few conjugate gradient

Table 4: GPU memory usage in MB and total runtime in seconds for Stage II and Stage III using F-AMOC and W-AMOC.

	Memory (MB)			Total Time (s)		
Equation	F-AMOC	W-AMOC	PINO correction	F-AMOC	W-AMOC	PINO correction
Darcy flow	103.71	7694.39	199.17	47.87	551.49	160.12
Heat eqn	130.04	2973.03	329.17	82.20	232.67	153.34
2D Poisson eqn	66.95	2078.98	280.54	66.92	195.12	156.47
3D Poisson eqn	76.63	14762.01	7136.81	352.40	10290.48	325.44
3D Helmholtz eqn	92.05	22010.61	10275.43	20.59	10677.19	350.25

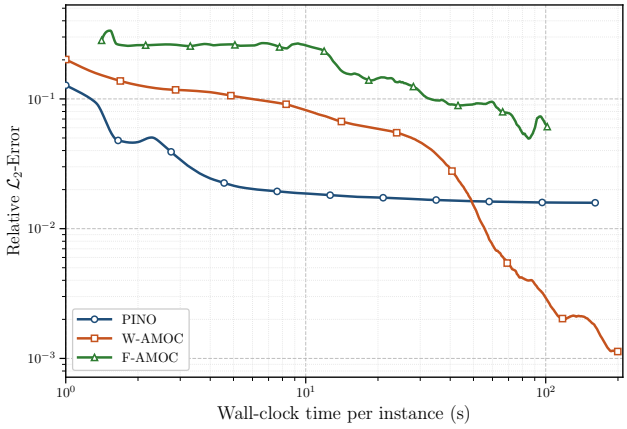

Figure 10: Relative $\mathcal{L}_2$-error versus wall-clock time for a single representative OOD instance of the 2D Poisson equation using PINO correction, W-AMOC, and F-AMOC.

iterations are enough. Stage III then works in the retained subspace, which has fewer parameters than the neural operator (Table 5). We use L-BFGS because this reduced problem is initialized from Stage II. L-BFGS converges at different rates along different Hessian eigendirections. This comes from the conditioning of the optimization problem and is separate from the representation-level spectral bias that AMOC targets, because each retained spectral component is an explicit optimization variable. SOAP or Adam could reduce the optimization cost further.

Table 5: Trainable parameters, retained basis functions M, and Stage III parameter block size.

	FNO/PINO	F-AMOC		W-AMOC	
Equation	Parameters	M	Block	M	Block
Darcy flow	1,188,385	200	797	3,831	19,156
Heat eqn	3,285,505	200	797	1,657	8,286
2D Poisson eqn	3,285,505	200	797	1,940	9,701
3D Poisson eqn†	131,080,993	200	1,001	14,430	14,430
3D Helmholtz eqn	131,081,025	200	1,001	15,072	105,505

†Basis functions are fixed and only coefficients are optimized.

C Hyperparameters

The hyperparameters for each stage of the proposed framework are specified separately for clarity and reproducibility. The configuration used in Stage I, corresponding to FNO pretraining for learning the coarse operator, is reported in Table 6. The hyperparameters for Stage II, which performs sparse spectral dominant basis extraction and initializes the corresponding spectral coefficients, are summarized in Table 7, where the dyadic dilation scales in each spatial dimension satisfy $j_i = 2^{J_i}$. The hyperparameters for Stage III, involving physics-constrained refinement, are provided in Table 8. In this stage, N_c denotes the number of interior collocation points, N_b the number of boundary points per side, and N_0 the number of initial condition points. The coefficients α, β, and γ control the relative weighting of the PDE residual, boundary condition loss, and initial condition loss, respectively.

Table 6: Hyperparameters used in Stage I

Parameter	Darcy	Heat	2D Poisson	3D Poisson	3D Helmholtz
Fourier layers	4	4	4	4	4
Channel width	32	32	32	32	32
Fourier modes	12	20	20	20	20
Input parameters	(x, y, k, f)	(x, t, ε)	(x, y, ε)	(x, y, z, f)	(x, y, z, k, f)
Input channels	4	3	3	4	5
Train grid	64×64	64×64	64×64	$64\times64\times64$	$64\times64\times64$
Test grid	64×64	64×64	64×64	$64\times64\times64$	$64\times64\times64$
Optimizer	Adam	Adam	Adam	Adam	Adam
Learning rate	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-3}
Weight decay	10^{-5}	10^{-5}	10^{-5}	10^{-5}	10^{-5}
Epochs	500	500	500	500	500
Batch size	10	10	10	10	10
Train samples	400	400	400	400	400
Spatial domain	$[0, 1]^2$	$[-1, 1]\times[0, 1]$	$[0, 1]^2$	$[0, 1]^3$	$[-1, 1]^3$
Test ID samples	10	10	10	10	10
Test OOD samples	10	10	10	10	10

Table 7: Hyperparameters used in Stage II.

Component	Parameter	2D Darcy	2D Heat	2D Poisson	3D Poisson	3D Helmholtz
W-AMOC	Scales J_x	$\{0,\dots,5\}$	$\{0,\dots,5\}$	$\{0,\dots,6\}$	$\{0,\dots,5\}$	$\{0,\dots,4\}$
	Scales J_y / J_t	$\{0,\dots,5\}$	$\{0,\dots,5\}$	$\{0,\dots,5\}$	$\{0,\dots,5\}$	$\{0,\dots,4\}$
	Scales J_z	—	—	—	$\{0,\dots,5\}$	$\{0,\dots,4\}$
	Padding Parameter a	0.3	0.2	0.2	0.3	0.3
	CG max iterations	1	1	5	1	2
	CG tolerance	10^{-1}	10^{-1}	10^{-2}	10^{-1}	10^{-1}
	Quantile level p	0.70	0.90	0.90	0.99	0.98
	Chunk size	128	256	256	128	256
	λ_{reg}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}
F-AMOC	Selected modes K	200	200	200	200	200

Table 8: Hyperparameters used in Stage III.

Component	Parameter	2D Darcy	2D Heat	2D Poisson	3D Poisson	3D Helmholtz
L-BFGS Optimizer	Learning rate	1.0	1.0	1.0	1.0	1.0
	Max iterations	5,000	5,000	5,000	5,000	5,000
	Max evaluations	10^{10}	10^{10}	10^{10}	10^{10}	10^{10}
	Gradient tolerance	10^{-10}	10^{-10}	10^{-10}	10^{-10}	10^{-10}
	Change tolerance	10^{-10}	10^{-10}	10^{-10}	10^{-10}	10^{-10}
	History size	50	50	50	50	50
Sampling Points	N_c	10,000	20,000	20,000	20,000	50,000
	N_b	3,000	1,000	1,000	2,000	5,000
	N_0	—	1,000	—	—	—
	No. of BC sides	4	2	4	6	6
	Test grid	128×128	128×128	128×128	$64\times64\times64$	$64\times64\times64$
Loss Weights	PDE weight α	1	0.01	0.005	0.01	0.01
	BC weight β	500	1	1	1	10
	IC weight γ	—	1	—	—	—

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction clearly state contributions (AMOC framework, multiscale correction, OOD gains) consistent with experimental and methodological results.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The paper explicitly discusses limitations in the "limitations and conclusion" section.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [N/A]

Justification: The paper does not present formal theoretical theorems or proofs but instead focuses on methodological development and empirical validation. Therefore, this item is not applicable.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: The paper provides detailed descriptions of datasets, PDE formulations, parameter distributions for ID and OOD settings, model architectures, training procedures, and evaluation metrics. These details are sufficient to reproduce the experiments. Additionally, the code and data were also given.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in

some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The paper provides access to code and data to reproduce the results.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: The paper specifies all relevant experimental details. Additional details are provided in the appendix tables for full transparency.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: The paper reports statistical variability through mean and standard deviation across multiple independent runs, as shown in Table 3 and Fig. 9.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The paper specifies the compute environment (NVIDIA RTX PRO 4500 GPU with CUDA 12.3) and provides details about training epochs, optimization procedures, and dataset sizes.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research adheres to the NeurIPS Code of Ethics. It involves numerical simulation and machine learning methods applied to PDEs, without human subjects, sensitive data, or ethical risks.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

Justification: This is a computational methods paper evaluated on synthetic PDE benchmarks with no human subjects, sensitive data, or deployable system; no foreseeable societal harm or misuse is anticipated at this stage.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: The work does not involve releasing high-risk models or datasets requiring misuse safeguards.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: The work builds on standard publicly available methods such as FNO and PDE benchmarks, which are properly cited. No proprietary datasets or restricted assets are used.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.

- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: The new assets, including the code and dataset used in this work, are provided via a link and are properly documented for reproducibility.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The research does not involve human subjects or crowdsourcing.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

Justification: The research does not involve human subjects and therefore does not require IRB approval.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: LLMs were not used as part of the core methodology. The use of LLMs was limited to writing or editing assistance and does not affect the scientific contributions.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.