1 Supplementary Notes: Cross-Domain PDE Benchmarks

1.1 Overview

This supplementary material provides a self-contained description of the three partial differential equation (PDE) benchmark families employed for cross-domain evaluation. The objective of these benchmarks is to assess whether physics-informed learning frameworks can acquire transferable, physically consistent representations that generalize across heterogeneous problem instances while strictly respecting the underlying governing equations.

Each benchmark family isolates a distinct and practically relevant source of distribution shift commonly encountered in scientific computing: (i) geometric variability of the computational domain, (ii) parametric heterogeneity in material properties and source terms, and (iii) structural variability of source configurations. Within each family, the mathematical form of the governing PDE is held fixed, and only the designated source of variability is modified across tasks.

Together, these benchmarks define a unified protocol for cross-domain generalization, enabling systematic evaluation beyond single-instance regression. All reference solutions are generated using high-fidelity numerical solvers with consistent discretization strategies to ensure fairness and comparability.

1.2 Supplementary Note S1: Geometry-Driven Linear Elasticity

This benchmark considers two-dimensional static linear elasticity under the small-deformation assumption and evaluates generalization across heterogeneous geometries. The primary goal is to examine whether learned representations preserve mechanical equilibrium and constitutive consistency when transferred across domains with distinct shapes and topologies.

Let $\Omega_i \subset \mathbb{R}^2$ denote the bounded Lipschitz domain associated with the i-th task. The displacement field $\mathbf{u}(x, y) = (u_x, u_y)$ satisfies the static equilibrium equations

$$\nabla \cdot \boldsymbol{\sigma} + \mathbf{F} = 0, \quad (x, y) \in \Omega_i, \tag{1}$$

where $\boldsymbol{\sigma}$ is the Cauchy stress tensor and $\mathbf{F}$ denotes a prescribed body force.

The constitutive relation follows isotropic linear elasticity,

$$\boldsymbol{\sigma} = \lambda \operatorname{tr}(\boldsymbol{\varepsilon})\mathbf{I} + 2\mu\, \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon}(\mathbf{u}) = \tfrac{1}{2}\left(\nabla \mathbf{u} + \nabla \mathbf{u}^{\top}\right), \tag{2}$$

where λ and μ are the Lamé parameters, which are fixed across all tasks.

Boundary conditions, material properties, and external forces are identical for all task instances. Task heterogeneity is introduced exclusively through variations in the domain geometry Ω_i, defining a geometry-driven cross-domain setting.

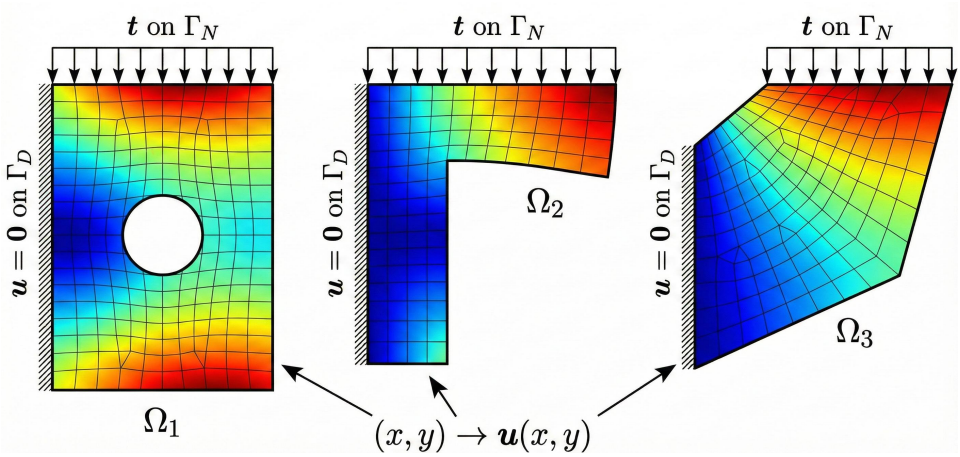

Figure S1: Representative domain geometries and corresponding displacement field distributions for the geometry-driven linear elasticity benchmark. All tasks share identical material parameters, boundary conditions, and loading, and differ only in domain geometry.

All reference solutions are computed using the finite element method implemented in FEniCS with sufficiently refined meshes to ensure numerical accuracy.

1.3 Supplementary Note S2: Parameter-Driven Steady-State Heat Conduction

This benchmark investigates steady-state heat conduction on a fixed spatial domain and evaluates generalization across strong parametric variations in material properties and internal heat sources.

Let $\Omega \subset \mathbb{R}^2$ denote a fixed bounded domain. The temperature field $T(x, y)$ satisfies the steady-state heat equation

$$-\nabla \cdot (k_i(x, y)\nabla T(x, y)) = Q_i(x, y), \quad (x, y) \in \Omega, \tag{3}$$

where $k_i(x, y)$ denotes the spatially varying thermal conductivity and $Q_i(x, y)$ represents the internal heat source associated with the i-th task.

The computational domain and boundary conditions are held fixed across all tasks. Task variability arises solely from different realizations of the conductivity field $k_i(x, y)$ and the source term $Q_i(x, y)$, defining a parameter-driven cross-domain setting.

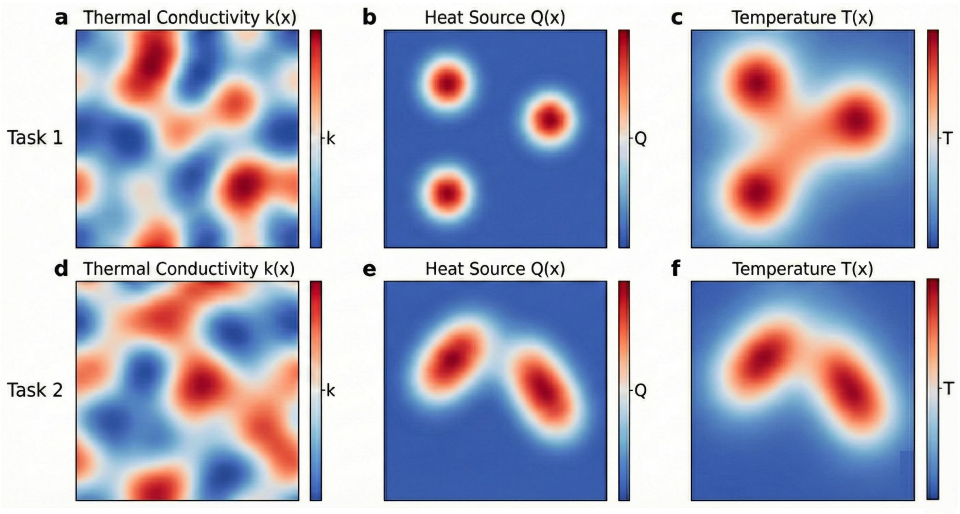

Figure S2: Characterization of parameter-driven steady-state heat conduction tasks. Rows illustrate Task 1 (**a**–**c**) and Task 2 (**d**–**f**) on a fixed domain. **a, d,** Heterogeneous thermal conductivity fields $k(\mathbf{x})$. **b, e,** Internal heat sources $Q(\mathbf{x})$ configured with localized Gaussian peaks. **c, f,** Corresponding steady-state temperature fields $T(\mathbf{x})$. The benchmark evaluates the model's ability to capture physical coherence across diverse material and source configurations.

All reference solutions are obtained using a consistent finite element discretization to ensure comparability across parameter realizations.

1.4 Supplementary Note S3: Source-Driven Steady-State Heat Conduction

This benchmark evaluates generalization across structurally diverse source configurations in steady-state heat conduction, isolating the spatial organization of heat sources as the sole source of task heterogeneity.

Let $\Omega \subset \mathbb{R}^3$ denote a fixed cubic domain. The steady-state temperature field $T(\mathbf{x})$ satisfies

$$-\nabla \cdot (k \nabla T(\mathbf{x})) = Q^{(i)}(\mathbf{x}), \quad \mathbf{x} \in \Omega, \tag{4}$$

where k is a constant thermal conductivity shared across all tasks.

Task variability is introduced exclusively through the source term $Q^{(i)}(\mathbf{x})$, which consists of multiple localized heat sources:

$$Q^{(i)}(\mathbf{x}) = \sum_{n=1}^{N_i} \exp\left(-a_{i,n} \left\|\mathbf{x} - \mathbf{x}_{i,n}\right\|^2\right), \tag{5}$$

where N_i denotes the number of sources, and $\mathbf{x}_{i,n}$ and $a_{i,n}$ represent their spatial locations and characteristic scales, respectively.

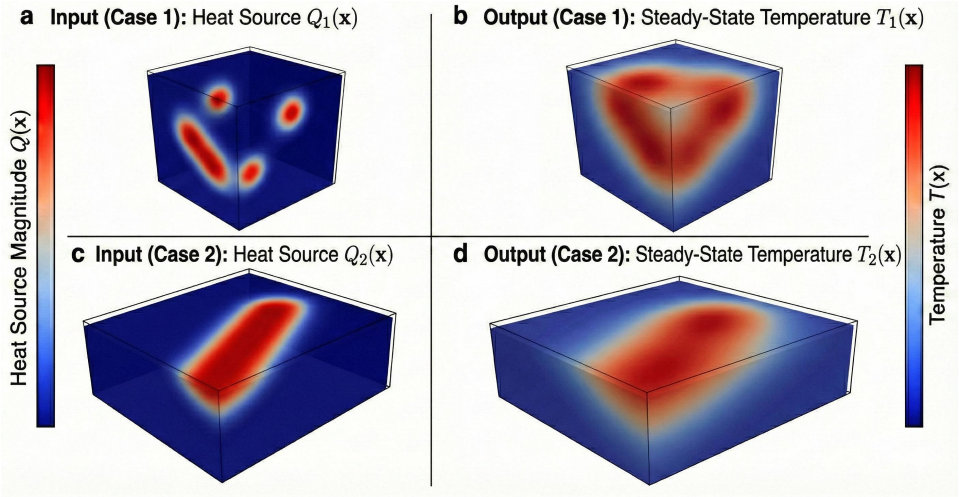

Figure S3: Representative source-driven heat conduction tasks with heterogeneous source configurations. Rows illustrate Case 1 (**a**–**b**) and Case 2 (**c**–**d**). **a, c,** Input source configurations $Q^{(i)}(\mathbf{x})$ showing the spatial layout of heat sources. **b, d,** Corresponding steady-state temperature fields $T(\mathbf{x})$. Distinct source structures give rise to markedly different solution manifolds, despite identical geometry and material properties.

All reference solutions are generated using high-resolution numerical solvers to accurately capture the resulting solution manifolds.

2 Supplementary Note: Hyperparameter Specification

2.1 Overview

This section details the network architectures, training protocols, and hyperparameter configurations employed to reproduce the results presented in the main text. Hyperparameter choices are guided by the principle of spectral–physical alignment, ensuring that the inductive biases of the neural network are consistent with the intrinsic structure of the PDE solution manifold. The selected configurations balance sufficient expressivity for operator approximation with regularization for stable transfer across heterogeneous geometric domains.

2.2 Data Preprocessing: Hessian Preconditioning

Proper conditioning of input and output spaces is critical for Physics-Informed Neural Networks (PINNs). Without normalization, multiscale physical quantities yield ill-conditioned Hessians, resulting in stiff optimization dynamics and convergence difficulties.

- **Domain Nondimensionalization (Spectral Alignment):** Neural networks exhibit a spectral bias, converging more rapidly on low-frequency components. To match this bias with the physical domain, spatial coordinates $\mathbf{x} \in \Omega$ are normalized by a characteristic length scale L_{ref} (e.g., hydraulic diameter):

$$\mathbf{x}^* = 2 \cdot \frac{\mathbf{x} - \mathbf{x}_{\min}}{L_{\text{ref}}} - 1 \in [-1, 1]^3. \tag{6}$$

  This mapping centers the domain within the active region of the tanh activation, maximizing gradient flow in early training stages.

- **Field Standardization (Isotropic Curvature):** Physical variables often span orders of magnitude (e.g., $p \sim 10^5$ Pa vs. $u \sim 10^{-1}$ m/s). Z-score standardization is applied using prototype task statistics:

$$\hat{y} = \frac{y - \mu_{\text{proto}}}{\sigma_{\text{proto}}}. \tag{7}$$

  This ensures approximately isotropic curvature in the loss landscape, preventing optimization from being dominated by high-magnitude quantities at the expense of transport features.

2.3 Architecture: Structural Disentanglement

The GD-PINNs framework explicitly separates *invariant physical laws* from *variant geometric realizations* using a split Multi-Layer Perceptron (MLP) backbone.

- **Universal Operator Encoder (Invariant Block):** Four hidden layers with 96-128 neurons each extract features corresponding to the differential operator $\mathcal{L}$, which is task-invariant. This depth provides sufficient capacity to approximate nonlinear operators while preserving transferability across tasks.
- **Geometric Manifold Adapter (Variant Block):** Three hidden layers with 96-128 neurons each map invariant features to geometry-specific boundary conditions. Limiting capacity prevents overfitting to local geometric variations and encourages reliance on the robust encoder features.
- **Graph Neural Network (Curvature Smoothing):** A 2-layer Graph Attention Network (GAT) with 4 heads is employed. Depth allows information propagation from two-hop neighborhoods to smooth local manifold curvature. LeakyReLU activations ($\alpha = 0.2$) mitigate vanishing gradients during message passing on sparse graphs.

2.4 Spectral–Physical Embedding: Manifold Coordinates

Task embedding vectors $\mathbf{z}_i$ serve as coordinates for the solution manifold. Total dimension $d = 150$ satisfies the Kolmogorov n-width criterion, ensuring bounded approximation error.

- **POD Modes ($d_{\text{POD}} = 10$):** Capture the energetic core via singular value decomposition (SVD) of solution snapshots. Top 10 modes retain $\geq 95\%$ cumulative energy:

$$K_{\text{POD}} = \min \left\{ k : \frac{\sum_{j=1}^{k} \sigma_j^2}{\sum_{j=1}^{N} \sigma_j^2} \geq 0.95 \right\}. \tag{8}$$

- **PSD Frequencies ($d_{\text{PSD}} = 60$):** Characterize geometric topology by retaining the top 20 dominant spatial frequencies per axis, effectively low-pass filtering high-frequency meshing artifacts.
- **FFT Components ($d_{\text{FFT}} = 80$):** Sample the fast Fourier transform of the solution fields to resolve multiscale dynamics up to the inertial sub-range, facilitating accurate transfer across flow regimes.

2.5 Graph Topology and Prototype Selection

The task graph $\mathcal{G}$ defines the pathways for knowledge diffusion. Prototype selection is determined by a composite centrality metric:

$$C_{\text{center}}(i) = \alpha\, C_{\text{eig}}(i) + (1-\alpha)\, C_{\text{deg}}(i). \tag{9}$$

The coefficient α is constrained to the interval $\alpha \in [0,1]$ to ensure a convex combination of the two topological metrics. This constraint is theoretically justified by the following considerations:

- Interpolation versus extrapolation: Constraining α to [0,1] ensures that the composite score remains strictly within the convex hull of the component metrics. Values outside this range would correspond to affine extrapolation, where one metric receives a negative weight (for instance, if $\alpha > 1$, then $1 - \alpha < 0$).
- Stability: A negative weight would imply penalizing a favorable topological feature, such as high local density or global connectivity, which is inconsistent with the goal of identifying influential nodes. This could lead to unstable prototype selection and increased sensitivity to outliers.
- Pareto optimality: The range [0,1] represents the valid Pareto frontier between two competing objectives—maximizing global information flow (C_{eig}) and maximizing local neighborhood coverage (C_{deg}).

Parameter specification and rationale:

- $\alpha = 0.65$ for SWRO industrial cases: This value favors eigenvector centrality (C_{eig}) because the industrial design space is fragmented and high-dimensional. A higher weight on global connectivity ensures that the prototype acts as a bridge capable of reaching distant geometric clusters that might otherwise remain isolated within the graph.
- $\alpha = 0.50$ for canonical benchmarks: This choice provides equal weighting between the two metrics. Benchmark tasks, such as elasticity and heat problems, are generated through uniform sampling of continuous parameters, forming a regular and convex manifold geometry. In such uniform settings, a balanced reference reflecting both local and global influence serves as an effective prototype.

2.6 Transfer Mechanism: Entropy-Controlled Diffusion

Knowledge propagation is implemented as anisotropic diffusion with temperature scaling:

$$w_{j\to i} \propto \exp(\langle \mathbf{h}_i, \mathbf{h}_j \rangle / \tau). \tag{10}$$

- **Temperature Annealing ($\tau : 2.0 \to 0.1$):** Early-stage high τ promotes uniform averaging (global exploration), while late-stage low τ enforces projection onto local tangent spaces for precise geometric adaptation, analogous to simulated annealing on the solution manifold.

2.7 Optimization: Boundary-Driven Uniqueness

Total loss is weighted as:

$$\mathcal{L} = \lambda_f \mathcal{L}_{\text{PDE}} + \lambda_b \mathcal{L}_{\text{BC}}. \tag{11}$$

- **Boundary Weighting:** $\lambda_b = 50$ (3D), $\lambda_f = 1$. The uniqueness of each task solution is entirely determined by boundary conditions; therefore, boundary adherence is prioritized.
- **Hybrid Optimization Protocol:** Adam optimizer (10,000 epochs) for stochastic exploration, followed by L-BFGS-B for precise convergence ($gtol = 10^{-5}$), ensuring scientific accuracy.

2.8 Hyperparameter Summary

Table S1 summarizes the configurations used in this study, ensuring reproducibility while maintaining spectral–physical consistency.

3 Supplementary Note: Theoretical Derivation of Manifold-Restricted Optimization

In this section, we provide a rigorous derivation of the initialization error bound presented in Eq. (23) of the main text. We show that the Graph-Guided Knowledge Diffusion in GD-PINNs acts as a locally valid tangent space interpolation, and that its error is fundamentally constrained by the curvature of the solution manifold and the spectral sampling density.

Table S1: Hyperparameter specifications for GD-PINNs. Configurations balance operator expressivity with geometric adaptability and enforce thermodynamic consistency in diffusion.

Category	**Parameter**	**Value / Specification**
Architecture	Encoder Structure (Invariant)	4 Layers × 96-128 Neurons
	Adapter Structure (Variant)	3 Layers × 96-128 Neurons
	Activation Function	tanh (C^∞ smooth)
	GNN Architecture	2 Layers, 4 Attention Heads
Embedding	Dimension Breakdown	$d_{\text{POD}} = 10$, $d_{\text{PSD}} = 60$, $d_{\text{FFT}} = 80$
	Total Dimension	150 (Kolmogorov n-width satisfied)
Graph Topology	Centrality Weight (α)	0.65 (SWRO) / 0.50 (Benchmark)
	Neighborhood Size (k)	5 (Nearest Neighbors)
Transfer	Diffusion Temp. (τ)	Annealed: $2.0 \rightarrow 0.1$
	Pre-training	Spectral-Graph Hard Transfer
Training	Optimizer Schedule	Adam ($1e^{-3}$) $\rightarrow$ L-BFGS ($1e^{-5}$)
	Loss Weights	$\lambda_f = 1.0$, $\lambda_b = 50$ (3D) / 20 (2D)
	Collocation Points (N_f)	20,000 (3D) / 5,000 (2D)
	Random Seed	42 (Fixed for reproducibility)

3.1 Geometric Setup and Assumptions

Let the parameter space of the neural network be $\Theta \cong \mathbb{R}^D$. We assume the existence of a smooth, low-dimensional submanifold $\mathcal{M}_\Theta \subset \Theta$ formed by the optimal parameters $\{\theta^*(\mu) : \mu \in \mathcal{P}\}$ corresponding to the PDE family.

Assumption 1 (Manifold Regularity): The submanifold $\mathcal{M}_\Theta$ is a compact Riemannian manifold with bounded sectional curvature. Let κ denote the upper bound of the second fundamental form (extrinsic curvature) of $\mathcal{M}_\Theta$ embedded in $\mathbb{R}^D$.

Assumption 2 (Spectral Isometry): The spectral-physical embedding $\psi : \mathcal{M}_{\text{PDE}} \to \mathbb{R}^d$ (constructed via PSD, FFT, POD) is approximately isometric in local neighborhoods. Consequently, the Euclidean distance in the embedding space approximates the geodesic distance on the parameter manifold:

$$\|\mathbf{z}_i - \mathbf{z}_j\|_2 \approx d_{\mathcal{M}}(\theta_i^*, \theta_j^*). \tag{12}$$

3.2 Graph Diffusion as Tangent Space Approximation

In GD-PINNs, the initialization θ_i^0 for a target task T_i is computed via attention-weighted aggregation:

$$\theta_i^0 = \sum_{j \in \mathcal{N}(i)} \alpha_{ij} \theta_j^*, \quad \text{with } \sum \alpha_{ij} = 1, \tag{13}$$

where $\mathcal{N}(i)$ denotes the set of spectral neighbors. Geometrically, this operation represents a convex combination of points lying on the curved manifold $\mathcal{M}_\Theta$. The resulting θ_i^0 typically lies on the secant plane connecting the neighbors rather than on the manifold itself.

We aim to bound the approximation error $E = \|\theta_i^0 - \theta_i^*\|_2$.

3.3 Derivation of the Error Bound

Consider the local geometry around the optimal parameter θ_i^* for the target task. Since $\mathcal{M}_\Theta$ is smooth, for any neighbor θ_j^* we can perform a Taylor expansion along the geodesic $\gamma_{ij}(t)$ connecting θ_i^* and θ_j^*.

Let $\mathbf{v}_{ij} \in T_{\theta_i^*}\mathcal{M}_\Theta$ denote the tangent vector corresponding to this geodesic, with $\|\mathbf{v}_{ij}\| = d_{\mathcal{M}}(\theta_i^*, \theta_j^*)$. In the ambient Euclidean space, the expansion is:

$$\theta_j^* = \theta_i^* + \mathbf{v}_{ij} + \frac{1}{2}\mathcal{H}_{\theta_i^*}(\mathbf{v}_{ij}, \mathbf{v}_{ij}) + O(\|\mathbf{v}_{ij}\|^3), \tag{14}$$

where $\mathcal{H}_{\theta_i^*}$ is the second fundamental form representing extrinsic curvature.

Substituting this into the aggregation formula:

$$\theta_i^0 = \sum_j \alpha_{ij} \left(\theta_i^* + \mathbf{v}_{ij} + \frac{1}{2}\mathcal{H}_{\theta_i^*}(\mathbf{v}_{ij}, \mathbf{v}_{ij}) + \dots \right) \tag{15}$$

$$= \theta_i^* + \sum_j \alpha_{ij}\mathbf{v}_{ij} + \frac{1}{2}\sum_j \alpha_{ij}\mathcal{H}_{\theta_i^*}(\mathbf{v}_{ij}, \mathbf{v}_{ij}) + \dots \tag{16}$$

Analysis of the terms:

1. Tangent Bias (First Order): The term $\sum \alpha_{ij}\mathbf{v}_{ij}$ represents the weighted mean of tangent vectors. The diffusion temperature τ in GD-PINNs is annealed to prioritize nearest neighbors symmetrically. Under dense, symmetric sampling or with appropriately centered attention weights, this term is minimized or approaches zero.

2. Curvature Error (Second Order): This term arises from the manifold geometry and cannot be eliminated by linear averaging. By the bound on the second fundamental form:

$$\left\| \mathcal{H}_{\theta_i^*}(\mathbf{v}_{ij}, \mathbf{v}_{ij}) \right\|_2 \leq \kappa \|\mathbf{v}_{ij}\|_2^2. \tag{17}$$

Let $\delta = \max_{j \in \mathcal{N}(i)} \|\mathbf{v}_{ij}\|$ denote the fill distance determined by the spectral graph connectivity. Then $\|\mathbf{v}_{ij}\| \leq \delta$.

Hence, the error is bounded by:

$$\|\theta_i^0 - \theta_i^*\|_2 \leq \left\| \sum \alpha_{ij}\mathbf{v}_{ij} \right\| + \frac{1}{2}\sum \alpha_{ij}\kappa\delta^2 + O(\delta^3). \tag{18}$$

Neglecting higher-order terms and assuming the tangent bias is negligible due to attention centering, we obtain the simplified bound:

$$\|\theta_i^0 - \theta_i^*\|_2 \leq \frac{1}{2}\kappa\delta^2 \equiv C(\kappa)O(\delta^2). \tag{19}$$

3.4 Connection to GD-PINNs Stages

This bound provides a theoretical justification for the dual-stage design of GD-PINNs:

Stage I (Universal Operator Encoder) reduces $C(\kappa)$: By fixing the encoder parameters θ_{enc} across tasks, the manifold is effectively flattened along the invariant operator dimensions. The variation of θ^* is then restricted to a lower-curvature submanifold defined by the geometric adapter θ_{adapt}, reducing the curvature constant κ.

Stage II (Manifold Interpolation) reduces δ: The spectral embedding $\mathbf{z}$ ensures that the graph selects neighbors that are physically close. Adaptive attention weights α_{ij} further reduce the effective fill distance δ, keeping the approximation within the quadratic region $O(\delta^2)$.