BUCKINGHAM π-INVARIANT TEST-TIME PROJECTION FOR ROBUST PDE SURROGATE MODELING

Seokki Lee[1]*, **Min-Chul Park**[2]*, **Giyong Hong**[1]*, **Changwook Jeong**[1]†
[1]Ulsan National Institute of Science and Technology (UNIST)
[2]CSE Team, Samsung Electronics
{seokki, e150great, changwook.jeong}@unist.ac.kr
m.c.park@samsung.com

ABSTRACT

PDE surrogate models such as FNO and PINN struggle to predict solutions across inputs with diverse physical units and scales, limiting their out-of-distribution (OOD) generalization. We propose a π-invariant test-time projection that aligns test inputs with the training distribution by solving a log-space least squares problem that preserves Buckingham π-invariants. For PDEs with multidimensional spatial fields, we use geometric representative π-values to compute distances and project inputs, overcoming degeneracy and singular points that limit prior π-methods. To accelerate projection, we cluster the training set into K clusters, reducing the complexity from $\mathcal{O}(MN)$ to $\mathcal{O}(KN)$ for the M training and N test samples. Across wide input scale ranges, tests on 2D thermal conduction and linear elasticity achieve MAE reduction of up to $\approx 91\%$ with minimal overhead. This training-free, model-agnostic method is expected to apply to more diverse PDE-based simulations.

1 INTRODUCTION

A central principle of physics is *dimensional homogeneity*: every valid physical law can be expressed in terms of dimensionless groups, known as *Buckingham π-invariants*. Once recast in terms of these invariants, a system's behavior remains unchanged under any rescaling of dimensional variables that preserves the values of the π-groups. For instance, flows in a narrow pipe and in a vast river are dynamically equivalent when their Reynolds number matches. This implies that certain apparent out-of-distribution (OOD) shifts, often regarded as a major challenge in physics-informed machine learning and neural operator models, may not constitute genuine distribution shifts at all, but merely *scale changes* that are physically equivalent under the Buckingham π framework.

Building on this observation, we propose a general, training-free, and model-agnostic test-time procedure, *π-invariant projection*. Given the input fields and parameters of a partial differential equation (PDE), we map each test sample into the neighborhood of the training distribution *while preserving its π-values*: specifically, we move the sample within its own *π-equivalence class* and project it onto the nearest training π-equivalence class. As a result, inputs are aligned according to their essential physical similarity rather than arbitrary scale. The framework is applicable to a broad class of PDEs wherever dimensional analysis identifies invariant groups, and integrates as a drop-in inference step for any surrogate model, including U-Nets and neural operators. To reduce computational cost, we replace exhaustive pairwise comparisons with a centroid scheme: clustering M training samples into K representatives yields $\mathcal{O}(KN)$ test-time complexity instead of $\mathcal{O}(MN)$ for N test instances. Across diverse physical simulations, π-invariant projection substantially improves OOD generalization, often approaching in-distribution accuracy.

*Equal contribution.
†Corresponding author.

2 RELATED WORKS

Fourier Neural Operator (FNO). The Fourier neural operator (FNO) is a resolution-agnostic surrogate that learns mappings in Fourier space, enabling resolution-independent predictions (Li et al., 2020). Numerous FNO variants followed (Bartolucci et al., 2023; Kossaifi et al., 2023; Rahman et al., 2022). However, truncating to a limited set of low-frequency modes can discard physically relevant high-frequency content; Physics-Informed Neural Operators (PINO) mitigate this by adding PDE-constrained losses to recover high-frequency behavior (Li et al., 2024). Overall, FNO excels at in-distribution interpolation but struggles to extrapolate OOD when inputs have disparate units and scales.

Dimensionless learning. Learning dimensionless π-groups reduces variables and enforces scale invariance, enabling scalar extrapolation with neural networks or SINDy (Bakarji et al., 2022; Xie et al., 2022; Oppenheimer et al., 2023; Gunaratnam et al., 2003; Brunton et al., 2016). However, most work focuses on scalars rather than spatial fields. For fields, naive Buckingham–π scaling has two failures: (i) numerator zeros collapse many inputs to zero; and (ii) near-zero denominators make π undefined or unbounded, destabilizing learning and analysis. These issues are rarely addressed, limiting direct use in 2D/3D. Recent efforts couple π-groups with nonlinear maps in turbulent flows (Fukami et al., 2024), but a systematic treatment of zero-set collapse and denominator blow-ups remains missing.

Generalization for the Out-of-distribution (OOD) data. Uncertainty methods estimate OOD error but seldom improve accuracy without fine-tuning (Gal & Ghahramani, 2016; Lakshminarayanan et al., 2017; Fuchsgruber et al., 2024; Angelopoulos et al., 2024). Test-time training/adaptation updates a pretrained model during inference to boost OOD performance (Gandelsman et al., 2022; Wang et al., 2020; Adachi et al., 2024), but adds optimization overhead and latency and remains rare for regression and spatial fields. A promising alternative is sample-wise, π-preserving test-time alignment without TTT, which is underexplored.

Our Contributions. This work addresses inefficiency and generalization challenges in PDE-based physical prediction caused by differences in input units and ranges, by introducing a test-time projection based on Buckingham π-invariance. The projection maps test samples near the training samples while preserving a governing dimensionless group, and a centroid-based scheme reduces computational complexity from $\mathcal{O}(MN)$ to $\mathcal{O}(KN)$. The module is training-free and model-agnostic, significantly improving OOD performance across various physical simulations, often approaching in-distribution accuracy. The key contributions are three-fold.

1. π-preserving test-time projection for spatial fields: an explicit, invariant input transform that maps each test sample toward the training distribution under the governing π, improving robustness to out-of-distribution shifts.
2. π-uniform strategy: uniformizes the sample-wise π distribution by tuning the dominant-scale input while others fixed, enabling balanced training coverage which can be a general methodology in dimensionless modeling.
3. centroid reduction for projection: replaces exhaustive pairwise comparisons with centroid representatives, reducing test-time complexity from $\mathcal{O}(MN)$ to $\mathcal{O}(KN)$ with negligible accuracy impact.

Although our proposed framework can be applied to a wide range of PDE analyses, we focused on thermal and stress problems. Additional examples of applying the PDE are shown in App. F. In Section 3, we introduce the background of Buckingham-π and define the dimensionless parameters in the thermal conduction and linear elasticity. Section 4, the method for model-agnostic π-invariant projection is explained. In addition, at the test-time, efficient method in terms of computation cost is introduced throughout the π-uniform strategy and K-means clustering. Section 5 describes the experimental procedure in detail, while Section 6 presents the corresponding results. Finally, Section 7 present a discussion on the experimental limitations and the conclusions.

3 BUCKINGHAM π THEOREM

The Buckingham π-theorem provides a framework for dimensional analysis, transforming physical variables into dimensionless π-groups to reduce complexity and ensure scale invariance. If we know the units of parameters in the given PDE, the π-groups are easily extracted by eliminating the units of denominator and numerator. For a system with n variables and m fundamental units (e.g., mass, length, time), it yields $n - m$ dimensionless groups (see App.A). By recasting the problem in terms of these invariants, seemingly distinct inputs with different scales can be recognized as physically equivalent, providing a theoretical foundation for robust OOD generalization. Below, we define representative π-groups for thermal conduction and stress simulation, used to motivate invariants and scaling rules.

Theorem 1 ((Buckingham π, log form)). *Given B base units, p dimensional variables $x \in \mathbb{R}^p_{>0}$ and a dimension matrix $D \in \mathbb{R}^{|B|\times p}$ (rows = base units, columns = variables), let $r = \operatorname{rank}(D)$. Then there exist $p - r$ independent dimensionless combinations ("π-groups"). If $\Phi \in \mathbb{R}^{p\times(p-r)}$ spans $\ker(D)$, then*

$$\log \Pi(x) = \Phi^\top \log x \quad \Longleftrightarrow \quad \Pi(x) = \exp(\Phi^\top \log x)$$

is a complete set of $p - r$ independent π-groups. (See App. A for details.)

Notation. $\Phi = [\phi^{(1)} \cdots \phi^{(p-r)}]$ stacks the null-space basis vectors; each column $\phi^{(\ell)}$ defines one dimensionless monomial (a π-group). The set of log-rescalings that preserve all π-values is $\ker(\Phi^\top) = \{\, v \in \mathbb{R}^p : \ \Phi^\top v = 0 \,\}$.

How scaling acts (log-space translation). A componentwise rescaling $x \mapsto x \odot \exp(v)$ becomes a translation $z = \log x \mapsto z + v$ in log space. If $v \in \ker(\Phi^\top)$, then $\Phi^\top(z+v) = \Phi^\top z$, i.e., π-values are unchanged. Thus each input z generates an affine π-*equivalence class* $z + \ker(\Phi^\top)$.

In short, the physics lives in the π-values: unit or scale changes are just π-preserving shifts in log space, so inputs related by such shifts are the same case.

Worked example (thermal). With base units (M, L, T, Θ) and variables $(k, q, \Delta T, L)$,

$$[k] = MLT^{-3}\Theta^{-1}, \quad [q] = ML^{-1}T^{-3}, \quad [\Delta T] = \Theta, \quad [L] = L.$$

, the dimension matrix D can be formed.

$$D = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 1 \\ -3 & -3 & 0 & 0 \\ -1 & 0 & 1 & 0 \end{bmatrix}, \qquad \text{columns: } \{k,\, q,\, \Delta T,\, L\}, \quad \text{rows: } \{M,\, L,\, T,\, \Theta\}. \tag{1}$$

In general, for a variable x_j with units $\prod_{b\in B} b^{D_{b,j}}$, the j-th column of D stores its exponents. The number of π groups defined by the Buckingham π theorem is $p - \operatorname{rank}(D)$. Because $\operatorname{rank}(D) = 3$, only one π value is defined by the theorem.

A dimensionless monomial corresponds to a null-space vector $\phi \in \ker(\mathrm{D})$, i.e., $D\phi = 0$. Stacking independent solutions gives a π-basis $\Phi = [\phi^{(1)} \cdots \phi^{(p-r)}]$ with $D\Phi = 0$, but in this case $p - r = 1$, so one vector suffices. A π-basis (null-space basis) is $\Phi = \left[\ \phi^{(1)}\ \right]$, where

$$\phi^{(1)} = \begin{bmatrix} -1 \\ 1 \\ -1 \\ 2 \end{bmatrix},$$

which produces

$$\pi_{\mathrm{th}} = k^{\phi_k} q^{\phi_q} (\Delta T)^{\phi_{\Delta T}} L^{\phi_L} = k^{-1} q^1 (\Delta T)^{-1} L^2 = \frac{q\, L^2}{k\, \Delta T}.$$

(M, L, T, Θ) are units of mass, length, time, and temperature, respectively. For the variables, k is a thermal conductivity field with its unit of $[W/(mK)]$, q is a volumetric heat source field with $[W/m^3]$, ΔT is a scalar value from the difference between maximum and minimum value of boundary conditions with $[K]$, and L is the grid length with $[m]$. The π_{th} is the constraint our projection enforces (thermal row of equation 12).

Worked example (elasticity). Elasticity proceeds analogously. With base units (M, L, T) and variables $(E, \sigma, f, L, \Delta u)$, the dimension matrix D can be formed.

$$D = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ -1 & -1 & -2 & 1 & 1 \\ -2 & -2 & -2 & 0 & 0 \end{bmatrix}, \qquad \text{columns: } \{E,\ \sigma,\ f,\ L, \Delta u\}, \quad \text{rows: } \{M,\ L,\ T\}. \tag{2}$$

For the variables, E is the Young's modulus field with its unit of $[Pa]$, σ is a stress field with $[Pa]$, f is a body force per volume field with $[N/m^3]$, L is the grid length with $[m]$, and Δu is a displacement value from the difference between maximum and minimum value of boundary conditions with $[m]$.

A π-basis (null-space basis) can be collected as

$$\Phi = \left[\ \phi^{(1)}\ \ \phi^{(2)}\ \ \phi^{(3)}\ \right], \qquad D\Phi = 0,$$

with columns

$$\phi^{(1)} = \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow \pi_1 = \sigma/E, \qquad \phi^{(2)} = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} \Rightarrow \pi_2 = fL/E, \qquad \phi^{(3)} = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} \Rightarrow \pi_3 = f\,\Delta u/E.$$

Details are in App. A for the task-specific π and the resulting linear log-constraint.

4 METHOD

We propose a training-free, model-agnostic π-invariant projection* to align OOD inputs with training data while preserving Buckingham π-groups (Section 3). The pipeline has three stages: (1) *π-preserving test-time projection* which projects each test sample toward the training data sample while preserving the π via dimension reduction, log-scale linearization, projection, and prediction; (2) *π-uniform strategy* that generate uniform $\log \pi$ distribution to cover the wide range of π group, improving the model prediction accuracy; (3) *Centroid reduction for projection*: combined with the π-uniform strategy, we run K-means clustering on log-space training features and use the K centroids as projection representatives.

4.1 PROBLEM DEFINITION

We are given a training dataset $\mathcal{D}_{\text{tr}} = \{(X_i, Y_i)\}_{i=1}^{M}$, where each input X_i consists of discretized input fields together with the associated physical parameters and discretization step, and each output Y_i is the corresponding discretized solution of a governing PDE. All variables and field values are dimensional; see Fig. 1. This dataset is used to train a chosen surrogate PDE model. Then, for a given test input $\tilde{X}$, our goal is to construct a transformed input $\tilde{X}^*$ that (i) preserves the Buckingham–π invariants of $\tilde{X}$ and (ii) lies close to the training distribution under a scale-appropriate metric.

More formally, let $\tilde{x} = \psi(\tilde{X})$ denote the vector of dimensional variables extracted from $\tilde{X}$, where ψ is a feature extractor. For algebraic equations, $\psi(\cdot) = \mathrm{Id}(\cdot)$. For differential equations, we need to extract the representative variables from $\tilde{X}$ using ψ, which will be discussed later. Then, the problem is defined as

$$\tilde{x}^* = \underset{\tilde{x}' \in [\psi(\tilde{X})]_\pi}{\operatorname{argmin}} \left[\operatorname{dist}(\tilde{x}', \{\psi(X_i)\}_{i=1}^{M})\right], \tag{3}$$

where $\operatorname{dist}(\cdot)$ is a chosen distance measure between a candidate point and the training set.

*The algorithm of the Buckingham π-invariant projection is shown in App. H.

4.2 DOMAIN PROFILE REDUCTION

For algebraic systems, π-groups can be extracted directly from variables by enforcing dimensional homogeneity in Theorem 1. However, applying Buckingham-π analysis directly to high-dimensional spatial fields in PDEs presents a fundamental challenge. A naive pixel-wise approach often leads to π-degeneracy, as shown in Fig. 1; if a field value is zero (e.g., $q \to 0$) or near-zero at a specific location, the resulting π-value may vanish or diverge, causing numerical instability.

To overcome this issue, we introduce a feature extractor $\psi : X \mapsto x$ that maps each discretized field to a finite set of characteristic variables. Among possible choices, we adopt the arithmetic mean of the field values. Unlike pixel-wise evaluation, this global statistic is robust to local zero values and outliers, ensuring non-zero representative scales that enable stable log-linear projection. For non-spatial inputs, these characteristic variables are used directly.

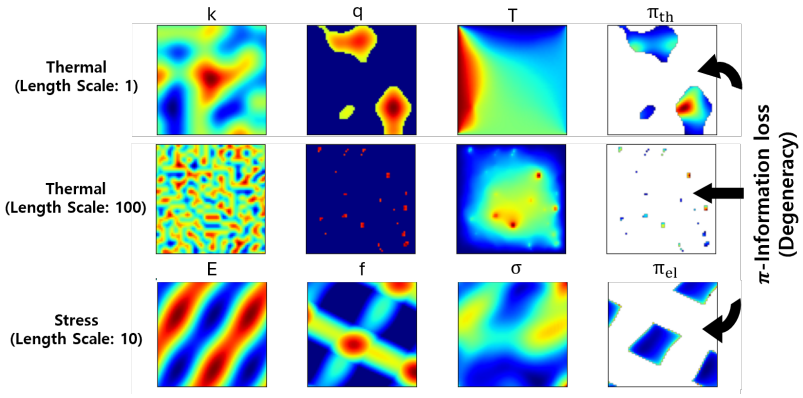

Figure 1: The PDE datasets for thermal and linear elasticity. With given distributions of k, q, E, f, the surrogate model predicts the fields of T and σ in each case (Thermal and Stress). Pixel-wise π values go to zero when $q \to 0$ and $f \to 0$, which is described as π-information loss.

4.3 OFFLINE PREPARATION: π-UNIFORM STRATEGY & CENTROID REDUCTION

The π-uniform strategy is adopted for enabling a balanced training coverage. It uniformizes the sample-wise π distribution in train samples by tuning the dominant scale input (e.g., q in thermal case) while others fixed. The dominant parameter is defined as the input parameter whose scale most directly controls the Buckingham π-group. Concretely, it can be done with SHAP analysis on group for each parameter. For example, the dominant parameter in thermal case shows the highest importance among whole parameter's contribution (48.7%). Once the dominant parameter is fixed, we first sample the input parameters from broad uniform ranges and compute their π-values. From this π-distribution, we define a target uniform distribution in π-space; each training sample adjusts only the dominant parameter (e.g., q for thermal). so that its π-value matches the target while other inputs fixed. By making K clusters in uniformized π distribution, the centers of K clusters represent whole distribution. In our approach, the distribution of $\log \pi$ is uniformized rather than that of π for enlarging the coverage. Naïvely, comparing N tests to all M training samples entails $\mathcal{O}(MN)$ pairwise evaluations of the inter-class residual in equation 3. To reduce this cost, we cluster training log-features with K-means and keep K centroids, cutting complexity to $\mathcal{O}(KN)$ while preserving projection accuracy (App. D; Sec. 5.5). (see App. C)

4.4 ONLINE TEST-TIME PROJECTION

The main goal is to find the optimal projection, i.e., the closest point between the *equivalence class* generated by the test point $\tilde{X}$ and *equivalence classes* generated by each train data (X_i, Y_i). Since the vector v that generates equivalence classes in log-space is simply a parallel translation, the space it spans is of the form $z + \ker(\Phi^\top)$. This implies that the entire space decomposes into two parts: (i) the component perpendicular to $\ker(\Phi^\top)$ (inter-class variation, i.e., the physical change), and

(ii) the component parallel to $\ker(\Phi^\top)$ (intra-class change, i.e., the scaling variation). Ultimately, we identify the closest training class to the test class in the inter-class direction, and then adjust the test sample within its own equivalence class to achieve a scale consistent with the chosen training samples.

More formally, setting $z_i = \log x_i$ and $\tilde{z} = \log \tilde{x}$, the optimization problem equation 3 becomes

$$\min_{\tilde{z}' \in \tilde{z} + \ker \Phi^\top} \operatorname{dist}\left(\tilde{z}', \{z_i\}_{i=1}^M\right) = \min_{i=1,\dots,M} \min_{v \in \ker \Phi^\top} \|z_i - (\tilde{z} + v)\|_2 = \min_{i=1,\dots,M} d_{\sim_\pi}([x_i], [\tilde{x}]), \quad (4)$$

where i indexes the training samples, $[\tilde{x}]_\pi$ and $[x_i]_\pi$ denote the *equivalence classes* of the test and i-th training samples, respectively. Thus, the problem reduces to finding the nearest training class in terms of *the quotient distance* between the fixed $[\tilde{x}]_\pi$ and the union of the training equivalence classes $\bigcup_{i=1}^M [x_i]_\pi$. See App. B for details on the used metric.

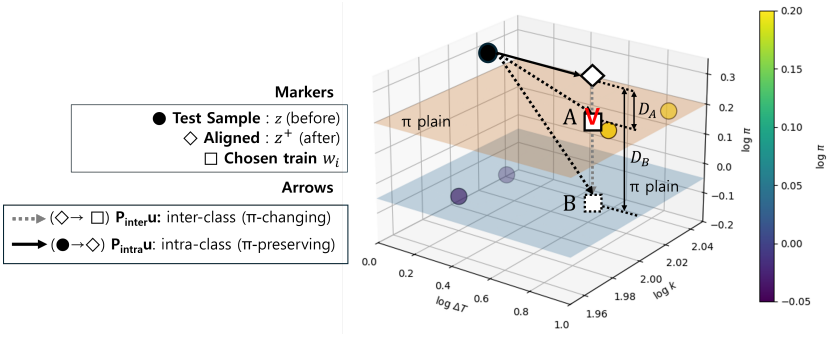

Figure 2: The schematic of π-invariant projection; the test sample (marked as circle) projects to the optimal train sample $x_A (\because D_A < D_B)$.

Step 1. Decomposition of optimal scaling vector

Let $v_i^t = z_i - \tilde{z} = \log(x_i/\tilde{x})$ in equation 4 denote the unconstrained log-scaling factor between x_i and $\tilde{x}$. Each v_i^t can be decomposed into *intra-class* (scaling) and *inter-class* (physics) parts using the projector onto $\ker(\Phi^\top)$:

$$\begin{aligned} P_\parallel &= I - \Phi(\Phi^\top\Phi)^{-1}\Phi^\top \quad \left(\text{or } P_\parallel = I - \Sigma^{-1}\Phi(\Phi^\top\Sigma^{-1}\Phi)^{-1}\Phi^\top \text{ with a weight } \Sigma\right), \\ v_i^t &= \underbrace{P_\parallel v_i^t}_{\text{intra-class (scales, } \pi\text{-preserving)}} + \underbrace{(I - P_\parallel)v_i^t}_{\text{inter-class (changes } \pi)} \end{aligned} \quad (5)$$

As an example, Figure 2 illustrates v_1^t and v_2^t, which are defined from the test sample $\tilde{x}$ and the two training candidates x_1 and x_2. For each i, choosing $v_i^* = P_\parallel v_i^t$ yields the quotient distance between $[x_i]$ and $[\tilde{x}]$, as the projection aligns the training and test samples in their respective affine spaces, isolating the class-wise irreducible component of the distance $(I - P_\parallel)v_i^t$.

Step 2. Optimum train sample to be projected

The optimum equivalence class to be projected has a minimum value of $\|v_i^t - v_i^*\|_2$ among whole training equivalence classes. The optimization problem can be represented as below:

$$i^* = \underset{i=1,\dots,M}{\operatorname{argmin}} \|v_i^t - v_i^*\|_2, \quad (6)$$

where i represents the equivalence class number to be projected. For example, in the Figure 2, $\tilde{x}_1^*$ is the optimum point because $D_A = \|v_1^t - v_1^*\|_2 < D_B = \|v_2^t - v_2^*\|_2$. After that, the optimally scaled test sample is finally given by

$$\tilde{x}^* = \exp(\tilde{z}^*) = \exp(\tilde{z} + v^*) = \tilde{x} \odot \exp(v^*).$$

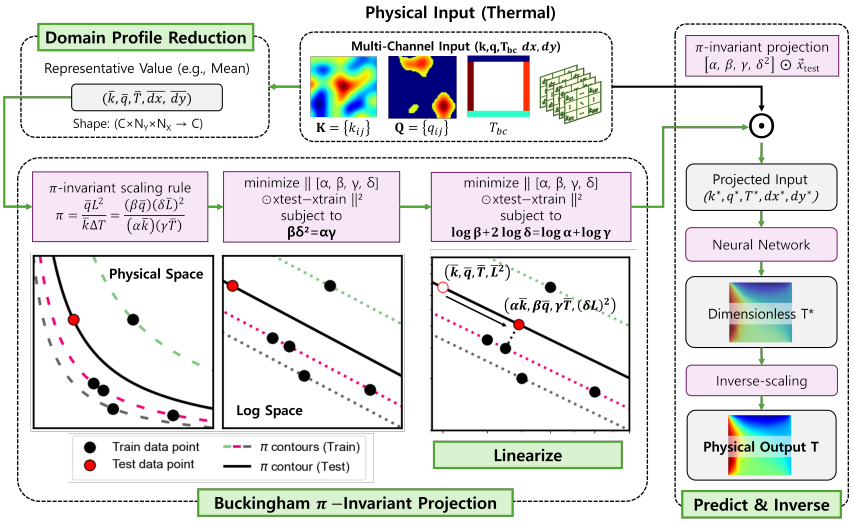

Figure 3: Flowchart of Buckingham π-Invariant Projection (Thermal case).

5 Experimental Setup

The experiment proceeds in the following order: Training (Sec. 5.1) $\rightarrow$ Domain Profile Reduction (Sec. 5.2) $\rightarrow$ π-invariant Projection (Sec. 5.3) $\rightarrow$ Prediction (Sec. 5.4) — to perform OOD data prediction (Figure 3). In addition, for more efficient π-invariant projection, the π-uniform strategy and centroid reduction method can also be employed (Section 5.5).

5.1 Training

Thermal conduction model We solve the steady 2D conduction problem with Dirichlet boundary conditions on T_{bc}:

$$-\nabla \cdot (k \nabla T) = q. \tag{7}$$

We use three surrogate architectures: a plain CNN, U-Net, and FNO. Training and test sets are generated by numerically solving equation 7 on uniform grids, but with *disjoint parameter ranges* to induce OOD shift. For example, $\log_{10} q \in [0,\ 7.5]$ for training and $\log_{10} q \in [7.5,\ 12]$ for testing. Distributions of k, q, T_{bc}, and dx are shown in Figure 4 (thermal panels). we encode this input into five channels (k, q, T_{bc}, dx, dy).

Linear elasticity model We consider plane stress/strain in 2D:

$$-\nabla \cdot \sigma = f, \quad \sigma = \mathbb{C}(E, \nu) : \varepsilon(u). \tag{8}$$

As in the thermal case, we evaluate CNN, U-Net, and FNO surrogates. Inputs are field maps E, ν, and f, the displacement boundary condition u_{bc}, and grid spacings dx, dy. We encode f and u_{bc} via components $(f_x, f_y, u_{bc}^x, u_{bc}^y)$, yielding eight input channels: $(E, f_x, f_y, u_{bc}^x, u_{bc}^y, \nu, dx, dy)$. Parameter distributions for (E, f, u_{bc}, dx) are shown in Figure 4 (elasticity panels).

5.2 Domain Profile Reduction

To represent global characteristics of the domain, arithmetic mean values of input fields are adopted. As defined in Section 4.2, fields of k, q (thermal) and E, f (elasticity) are summarized to representative values $\bar{k}$, $\bar{q}$, $\bar{E}$, $\bar{f}$ respectively. However, representive f ($\bar{f}$) may be extracted in a slightly different manner, because it is represented as two channels corresponding to f_x, f_y components. Therefore, in this case we first compute the means $\bar{f}_x$, $\bar{f}_y$ of f_x, f_y, respectively, and then compute $\bar{f}$ using the L_2-norms of $\bar{f}_x$, $\bar{f}_y$ ($\bar{f} = \sqrt{\bar{f}_x{}^2 + \bar{f}_y{}^2}$). we set $dx = dy = L$ unless otherwise stated. For quantities that have zero values are not considered. The resulting feature vectors $\tilde{x}$ are shown in the pipeline (Figure 3).

5.3 π-INVARIANT PROJECTION

Prior to projection, the scaling process is essential in projection, enabling the surrogate model to relocate OOD onto an interpolative (in-distribution) scale and reliable predictions. For the thermal case, the scaling coefficients α, β, γ and δ are used to scale the parameters while preserving the π_{th}. Log-scale linearization is applied for efficient calculation (Section 3).

$$\bar{k}' = \alpha\bar{k}, \quad \bar{q}' = \beta\bar{q}, \quad \Delta T' = \gamma\Delta T, \quad L' = \delta L \tag{9}$$

For preserving the π_{th}, scaling coefficients should follow the constraint.

$$\pi_{\text{th}} = \frac{\bar{q}'L'^2}{\bar{k}'\Delta T'} = \frac{(\beta\bar{q})(\delta L)^2}{(\alpha\bar{k})(\gamma\Delta T)} = \frac{\bar{q}L^2}{\bar{k}\Delta T} \tag{10}$$

Likewise, scaling coefficients are applied to each input in stress case as below:

$$\begin{aligned} \bar{E}' = \alpha\bar{E}, \quad \bar{f}' = \beta\bar{f}, \quad L' = \gamma L \\ \pi_{\text{el}} = \frac{\bar{f}'L'}{\bar{E}'} = \frac{(\beta\bar{f})(\gamma L)}{(\alpha\bar{E})} = \frac{\bar{f}L}{\bar{E}} \end{aligned} \tag{11}$$

Log formats are used to equations 25, 11 linearly transform scales of parameters.

$$\begin{cases} \log\beta + 2\log\delta - \log\alpha - \log\gamma = 0 & \text{(thermal)} \\ \log\beta + \log\gamma - \log\alpha = 0 & \text{(elasticity)} \end{cases} \tag{12}$$

which preserves π_{th}, π_{el} in Appendix A.1, A.2. After that, inputs $[k, q, \Delta T, L]$ of given test sample $\tilde{x}$ are scaled by $\exp(v^*) = [\alpha, \beta, \gamma, \delta]$, and moved to the optimal projection point $\tilde{x}^*$ having inputs $\tilde{x}^* = \tilde{x} \odot \exp(v^*) = [\alpha k, \beta q, \gamma\Delta T, \delta L]$ from calculating the nearest class with π-preserving correction described in Section 4.4. The same rule is applied in the calculation of the elasticity.

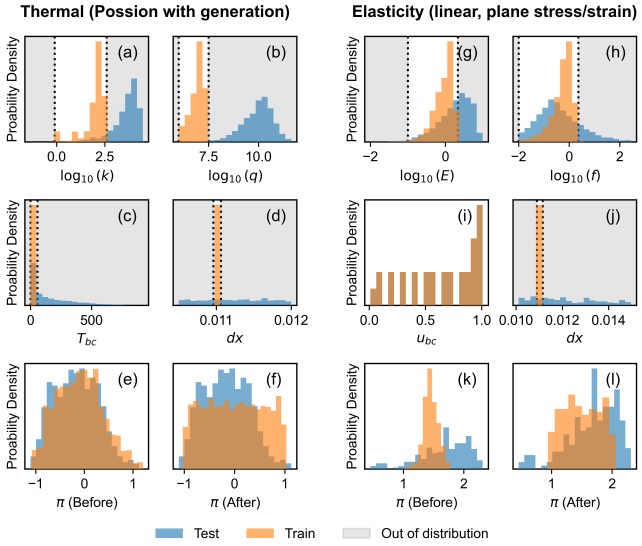

Figure 4: Histogram comparison for train (blue) and test (orange). Thermal (a–f): (a) $\log_{10} k$, (b) $\log_{10} q$, (c) T_{bc}, (d) dx; (e) $\log_{10} \pi_{\text{th}}$ before and (f) after π-uniform sampling. Elasticity (g-l): (g) $\log_{10} E$, (h) $\log_{10} \|f\|$, (i) u_{bc}, (j) dx, (k,l) $\log_{10} \pi_{\text{el}}$ (before vs. after).

5.4 PREDICTION PROCESS (SCALING, INVERSE SCALING)

After computing the optimal v^*, we rescale the test input channel-wise $\tilde{X}^* = \tilde{X} \odot \exp(v^*)$, i.e.,

$$\tilde{X}^* = \begin{cases} [\alpha k, \beta q, \gamma T_{bc}, \delta dx, \delta dy] & \text{(thermal)} \\ [\alpha E, \beta f_x, \beta f_y, \gamma u^x_{bc}, \gamma u^y_{bc}, \nu, \delta dx, \delta dy] & \text{(elasticity)} \end{cases} \tag{13}$$

Feeding $\tilde{X}^*$ to the surrogate model predicts T_{scaled}, and we invert the temperature scaling using the ΔT scale factor γ: $T = T_{\min} + \gamma^{-1}\big(T_{\text{scaled}} - T_{\min}\big)$, where $T_{\min}$ is the minimum Dirichlet boundary temperature. (Analogous inverse maps apply for elasticity outputs.)

5.5 CENTROID REDUCTION FOR PROJECTION

In calculation the nearest class in Section 5.3, We apply a π-uniform sampling strategy (Sec.4.3) to obtain an approximately uniform distribution of $\log \pi$ (Figure 4 (e,f), (k,l)). Prior to applying the strategy, the $\log \pi$ distribution is one-sided and skewed; afterwards it becomes approximately uniform, indicating that the training space covers a broader range of π values. For thermal we vary q to target desired π_{th} values; for elasticity we analogously vary f for π_{el}. After π-uniform sampling, we run K-means on the training log-features z to obtain $K \in \{1, \ldots, 10\}$ centroids. Because K-means is stochastic, we average results over 10 seeds (for both clustering and test sets). We compare *Baseline* (nearest over all training samples), *Clustered* (nearest over K centroids), and *Random* (nearest over K random training samples) using MAE, RMSE, and wall-clock. Timing excludes surrogate forward passes and reports only the projection stage (candidate selection + solving equation 3).

6 RESULTS AND DISCUSSION (THERMAL AND ELASTICITY)

Figure 5 compares *Top-3 best* and *Top-3 worst* OOD cases for thermal and elasticity. Across both, π-projection substantially improves fidelity, with the largest gains in the worst cases where the raw surrogate fails to extrapolate. Table 1 reports MAE/MSE and projection time for each architecture[†]. With the Buckingham-π projection, both thermal and stress simulations achieve lower MAE and RMSE errors, but the inference time increases. However, applying the cluster algorithm with π significantly reduce the inference time, while achieving a level of accuracy comparable to that obtained when computing distances over all training data.

The centroid reduction algorithm as described in 5.5 has demonstrated the efficiency in terms of computation cost when predicting the temperature distribution without a significant degradation in accuracy. From Figure 6, it is shown that the MAE tends to converge with the baseline projection as the number of candidates (cluster centers and random samples) increases while making a gap with random projection. The clustered projection matches Baseline MAE within statistical noise while reducing projection time by $\sim 100\times$ (Figure 6(c)).

Handling π-degeneracy. When π_{th} is ill-defined or ill-conditioned (e.g., $q \approx 0$ or $\Delta T \approx 0$), we drop the explicit π constraint and solve the same log-space alignment on the non-degenerate channels (e.g., $k, \Delta T, dx, dy$ in thermal). This removes global scale mismatch while preserving spatial heterogeneity, which empirically restores OOD accuracy. When valid π-groups are available (e.g., advection–diffusion, Navier–Stokes), the same log-affine machinery applies with the corresponding constraint.

Table 1: Test scores on thermal and elasticity. The best scores are bolded.

Method	**Thermal**			**Stress**		
	MAE	RMSE	Time	MAE	RMSE	Time
CNN	8.43±0.29	9.99±0.33	-	0.96±0.08	1.17±0.10	-
CNN + Pairwise Projection	2.63±0.04	3.24±0.05	100.31±2.06	**0.53±0.04**	**0.71±0.08**	73.58±0.59
CNN + π-uniform + 10-Randoms	1.85±0.14	2.30±0.17	**1.75±0.02**	0.62±0.05	0.86±0.08	**1.31±0.04**
CNN + π-uniform + 10-Centroids	**1.79±0.08**	**2.23±0.09**	1.80±0.02	0.60±0.05	0.84±0.08	1.36±0.04
U-Net	13.60±0.31	15.29±0.35	-	0.81±0.08	0.99±0.10	-
U-Net + Pairwise Projection	1.75±0.01	2.31±0.02	99.12±3.90	**0.17±0.03**	**0.28±0.08**	94.91±1.70
U-Net + π-uniform + 10-Randoms	1.20±0.11	1.56±0.13	**2.25±0.08**	0.23±0.03	0.40±0.08	**1.50±0.05**
U-Net + π-uniform + 10-Centroids	**1.18±0.08**	**1.53±0.10**	2.31±0.07	0.22±0.03	0.39±0.08	1.58±0.07
FNO	9.88±0.23	11.43±0.26	-	3.20±0.84	4.19±1.03	-
FNO + Pairwise Projection	1.38±0.02	1.74±0.02	151.44±3.16	**0.28±0.03**	**0.42±0.08**	94.02±1.10
FNO + π-uniform + 10-Randoms	1.29±0.12	1.65±0.14	**2.24±0.04**	0.34±0.03	0.54±0.08	**1.45±0.03**
FNO + π-uniform + 10-Centroids	**1.25±0.10**	**1.60±0.11**	2.31±0.06	0.33±0.03	0.53±0.08	1.54±0.04

7 LIMITATIONS AND CONCLUSION

Limitations. Our π-invariant projection assumes PDE-governed behavior and may underperform in hybrid settings with empirical components. Using representative statistics (e.g., arithmetic means

[†]We follow the authors' public implementation and adapt it to image-wise outputs; details in App. E.

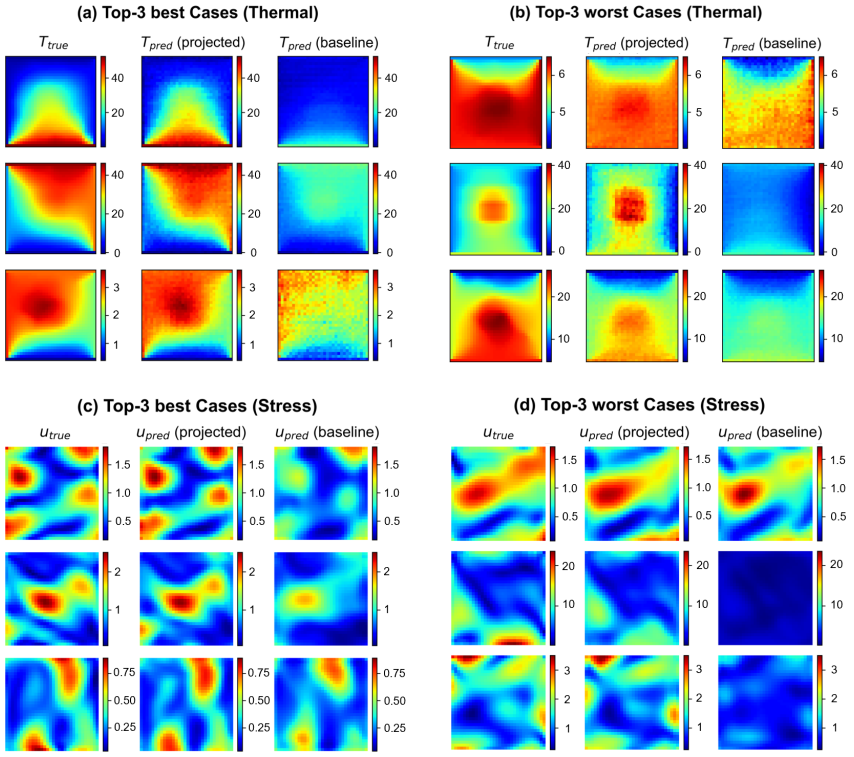

Figure 5: Comparison of true and predicted fields for *Top-3 best* (a) and *Top-3 worst* (b) cases in thermal (temperature) and elasticity (stress), under raw input vs. Buckingham π-invariant projection.

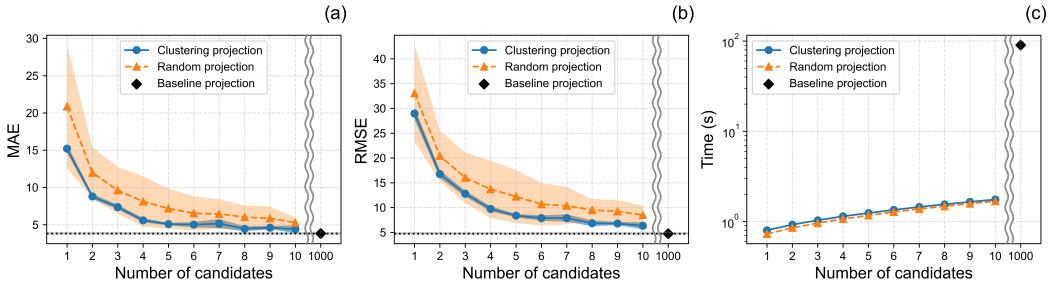

Figure 6: Performance comparison on (a) MAE, (b) RMSE, and (c) Time cost between clustering/random/baseline projection methods with 10 different test sets.

$\bar{k}$, $\bar{q}$) can blur highly irregular spatial patterns. While π-uniform sampling widens scale coverage, extreme OOD beyond the trained π range still degrades accuracy (see App. G).

Conclusion. We introduce a compact, training-free, and model-agnostic π-invariant test-time projection that enforces physical similarity via Buckingham-π and aligns OOD inputs to the training manifold through a tiny log-space least-squares scaling. This explicit log-scale transformation enables accurate predictions even under extreme out-of-distribution shifts. A centroid-based clustering variant accelerates inference, reducing complexity from $O(MN)$ to $O(KN)$ with negligible accuracy loss, and outperforms traditional test-time training in speed while maintaining high performance. On 2D thermal conduction and linear elasticity, the method reduces MAE by up to $\approx 91\%$ with minimal overhead. Future work includes uncertainty-aware projections, extensions to transient and convective PDEs (e.g., advection–diffusion, Navier–Stokes), and hybrid settings that incorporate empirical components.

ACKNOWLEDGMENTS

Changwook Jeong acknowledges support by the National Research Foundation of Korea (NRF), funded by the Ministry of Science and ICT (MSIT) (RS-2023-00257666, RS-2024-00458251, RS-2024-00408180); by the Ministry of Trade, Industry and Energy (MOTIE) (RS-2023-00231956, P0023703); by the Institute for Information & Communications Technology Planning & Evaluation (IITP), funded by MSIT (RS-2020-II201336, AIGS); by the research project fund of UNIST (1.250004.01; 1.220104.01); and by Samsung Electronics.

REFERENCES

Kazuki Adachi, Shin'ya Yamaguchi, Atsutoshi Kumagai, and Tomoki Hamagami. Test-time adaptation for regression by subspace alignment. *arXiv preprint arXiv:2410.03263*, 2024.

Anastasios N Angelopoulos, Rina Foygel Barber, and Stephen Bates. Theoretical foundations of conformal prediction. *arXiv preprint arXiv:2411.11824*, 2024.

Joseph Bakarji, Jared Callaham, Steven L Brunton, and J Nathan Kutz. Dimensionally consistent learning with buckingham pi. *Nature Computational Science*, 2022.

Francesca Bartolucci, Emmanuel de Bezenac, Bogdan Raonic, Roberto Molinaro, Siddhartha Mishra, and Rima Alaifari. Representation equivalent neural operators: a framework for alias-free operator learning. *Advances in Neural Information Processing Systems*, 36:69661–69672, 2023.

Steven L Brunton, Joshua L Proctor, and Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 2016.

Dominik Fuchsgruber, Tom Wollschläger, and Stephan Günnemann. Energy-based epistemic uncertainty for graph neural networks. *Advances in Neural Information Processing Systems*, 37: 34378–34428, 2024.

Kai Fukami, Susumu Goto, and Kunihiko Taira. Data-driven nonlinear turbulent flow scaling with buckingham pi variables. *Journal of Fluid Mechanics*, 2024.

Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pp. 1050–1059. PMLR, 2016.

Yossi Gandelsman, Yu Sun, Xinlei Chen, and Alexei Efros. Test-time training with masked autoencoders. *Advances in Neural Information Processing Systems*, 35:29374–29385, 2022.

David J Gunaratnam, Taivas Degroff, and John S Gero. Improving neural network models of physical systems through dimensional analysis. *Applied Soft Computing*, 2(4):283–296, 2003.

Jean Kossaifi, Nikola Kovachki, Kamyar Azizzadenesheli, and Anima Anandkumar. Multi-grid tensorized fourier neural operator for high-resolution pdes. *arXiv preprint arXiv:2310.00120*, 2023.

Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems*, 30, 2017.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.

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.

Michael W Oppenheimer, David B Doman, and Justin D Merrick. Multi-scale physics-informed machine learning using the buckingham pi theorem. *Journal of Computational Physics*, 2023.

Md Ashiqur Rahman, Zachary E Ross, and Kamyar Azizzadenesheli. U-no: U-shaped neural operators. *arXiv preprint arXiv:2204.11127*, 2022.

Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. *arXiv preprint arXiv:2006.10726*, 2020.

Xiaoyu Xie, Arash Samaei, Jiachen Guo, Wing Kam Liu, and Zhengtao Gan. Data-driven discovery of dimensionless numbers and governing laws from scarce measurements. *Nature communications*, 2022.

A Dimensional Analysis via Matrix Null Space

By following the processes for each task, Buckingham π groups can be defined.

A.1 Thermal conduction

Let the base dimensions be $B = \{M, L, T, \Theta\}$ (i.e. mass M, length L, time T, and temperature Θ). We encode units as exponent vectors x over B. For example, the unit of thermal conductivity (k) $[\mathrm{W}/(\mathrm{m}\cdot\mathrm{K})]$ is expressed as $x_k = [M^1 L^1 T^{-3} \Theta^{-1}]$, and taking a transpose, it is transformed into $x_k^\top = [1,\ 1,\ -3,\ -1]^\top$. Stacking the columns $x_k^\top, x_q^\top, x_{\Delta T}^\top, x_L^\top$ for the variables $[k, q, \Delta T, L]$ forms the dimension matrix $D \in \mathbb{R}^{|B|\times p}$ where p is a total number of input and output parameters. The units and parameters are :
k (thermal conductivity, $[\mathrm{W}/(\mathrm{m}\cdot\mathrm{K})]$), q (volumetric heat source, $[\mathrm{W}/\mathrm{m}^3]$), L (grid length, $[\mathrm{m}]$), and T (temperature, $[\mathrm{K}]$).

$$D = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 1 \\ -3 & -3 & 0 & 0 \\ -1 & 0 & 1 & 0 \end{bmatrix}, \qquad \text{columns: } \{k,\ q,\ \Delta T,\ L\}, \quad \text{rows: } \{M,\ L,\ T,\ \Theta\}. \tag{14}$$

In general, for a variable x_j with units $\prod_{b\in B} b^{D_{b,j}}$, the j-th column of D stores its exponents. The number of π groups defined by the Buckingham π theorem is $p - \mathrm{rank}(D)$. Because $\mathrm{rank}(D) = 3$, only one π value is defined by the theorem.

A dimensionless monomial corresponds to a null-space vector $\phi \in \ker(\mathrm{D})$, i.e., $D\phi = 0$. Stacking independent solutions gives a π-basis $\Phi = [\phi^{(1)} \cdots \phi^{(p-r)}]$ with $D\Phi = 0$, but in this case $p - r = 1$, so one vector suffices. A π-basis (null-space basis) is $\Phi = \left[\ \phi^{(1)}\ \right]$, where

$$\phi^{(1)} = \begin{bmatrix} -1 \\ 1 \\ -1 \\ 2 \end{bmatrix},$$

which produces

$$\pi_{\mathrm{th}} = k^{\phi_k} q^{\phi_q} (\Delta T)^{\phi_{\Delta T}} L^{\phi_L} = k^{-1} q^1 (\Delta T)^{-1} L^2 = \frac{q\, L^2}{k\, \Delta T}.$$

Unit check:

$$[\pi_{\mathrm{th}}] = \frac{(\mathrm{W}/\mathrm{m}^3)\cdot(\mathrm{m}^2)}{(\mathrm{W}/(\mathrm{m}\cdot\mathrm{K}))\cdot(\mathrm{K})} = \frac{\mathrm{W}/\mathrm{m}}{\mathrm{W}/\mathrm{m}} = 1,$$

so π_{th} is dimensionless (any nonzero scalar multiple of ϕ yields the same π up to a constant power).

A.2 Linear elasticity

We use base dimensions $B = \{M, L, T\}$ (no temperature). Let the variables be $(E, \sigma, f, L, \Delta u)$ with units $[E] = [\,\sigma\,] = M\,L^{-1}T^{-2}$, $[f] = M\,L^{-2}T^{-2}$, and $[L] = [\Delta u] = L$. The units and parameters are:
E (Young's modulus, $[\mathrm{Pa}]$), f (body force per volume, $[\mathrm{N}/\mathrm{m}^3]$), L (grid length, $[\mathrm{m}]$), u (displacement, $[\mathrm{m}]$), and σ (stress, $[\mathrm{Pa}]$).
The corresponding dimension matrix (rows M, L, T; columns as ordered) is

$$D = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ -1 & -1 & -2 & 1 & 1 \\ -2 & -2 & -2 & 0 & 0 \end{bmatrix}.$$

Here $p = 5$ and $\mathrm{rank}(D) = 2$, so the nullity is $p - \mathrm{rank}(D) = 3$, i.e., *three* independent π-groups.

A π-basis (null-space basis) is collected as

$$\Phi = \left[\ \phi^{(1)}\ \ \phi^{(2)}\ \ \phi^{(3)}\ \right], \qquad D\Phi = 0,$$

with columns

$$\phi^{(1)} = \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow \pi_1 = \sigma/E, \qquad \phi^{(2)} = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} \Rightarrow \pi_2 = fL/E, \qquad \phi^{(3)} = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} \Rightarrow \pi_3 = f\,\Delta u/E.$$

(If grid spacings are anisotropic, replace L by the arithmetic mean $L_{\text{eff}} = \sqrt{dx\,dy}$; equivalently, use $\frac{1}{2}(\log dx + \log dy)$ in the log constraints.)

A.2.1 MULTI π-INVARIANT PROJECTION

At test-time we enforce only *input-side* π-groups to avoid using unknown outputs. Thus we exclude $\pi_1 = \sigma/E$ and enforce $\pi_2 = fL/E$; if Δu is provided as an input, we may additionally enforce $\pi_3 = f\,\Delta u/E$ for a tighter projection. Default: if Δu is unavailable, enforce only $\pi_2 = fL/E$.

Let the input-side log-scales be $\mathbf{v} = [\log\alpha, \log\beta, \log\gamma, \log\delta]^\top$ acting on $E, f, L, \Delta u$ respectively, i.e.,

$$E' = \alpha E, \qquad f' = \beta f, \qquad L' = \gamma L, \qquad \Delta u' = \delta\,\Delta u.$$

Then the enforced π-constraints become linear equations in $\mathbf{v}$:

$$\underbrace{\begin{bmatrix} -1 & 1 & 1 & 0 \\ -1 & 1 & 0 & 1 \end{bmatrix}}_{C} \mathbf{v} = \mathbf{0} \quad \Longleftrightarrow \quad \begin{cases} \log\beta + \log\gamma - \log\alpha = 0 & (\pi_2 : fL/E), \\ \log\beta + \log\delta - \log\alpha = 0 & (\pi_3 : f\Delta u/E). \end{cases}$$

Given test log-features $\tilde{z}$ and a candidate train z_i, define $v_i^t = z_i - \tilde{z}$. Project v_i^t onto the π-preserving subspace using either

$$\text{(i) projector onto } \ker(C): \quad P_\parallel = I - C^\top (CC^\top)^{-1} C \quad \Rightarrow \quad v^\star = P_\parallel\, v_{i^\star}^t$$

Or equivalently

$$\text{(ii) KKT solve for } \min_{\mathbf{v}} \|\,\mathbf{v} - v_i^t\,\|_2^2 \text{ s.t. } C\mathbf{v} = 0.$$

Step 1 (nearest class under multiple π's).

$$i^\star \;=\; \arg\min_i \left\| (I - P_\parallel)\, v_i^t \right\|_2 . \tag{15}$$

Step 2 (π-preserving correction). Apply $v^\star = P_\parallel\, v_{i^\star}^t$ and rescale channels by $(\alpha, \beta, \gamma, \delta) = \exp(\mathbf{v}^\star)$ to obtain the projected input. Run the surrogate on the projected input and inverse-scale the outputs (details in App. B).

Notes. (i) If only one of $\{L, \Delta u\}$ is present, the system reduces to the single-constraint case (enforce fL/E or $f\Delta u/E$).
(ii) Under anisotropic discretization, replace the L-row in C by $\frac{1}{2}[\cdots,\ \log dx + \log dy\ ]$ (i.e., use $L_{\text{eff}} = \sqrt{dx\,dy}$).
(iii) If a constraint becomes ill-conditioned (e.g., $f \approx 0$), drop that row of C and project with the remaining valid constraints (degeneracy fallback).

B INPUT-SPACE L2 METRIC ON LOG-SCALE

Definition 1 (Logarithmic distance)**.** *For $x, y \in \mathbb{R}^p_{>0}$ with logarithmic coordinates $z = \log x$ and $w = \log y$, define the distance*

$$d(x, y) = \|\log x - \log y\|_2 = \|z - w\|_2.$$

Then, d is a metric on $\mathbb{R}^p_{>0}$, because $x \mapsto \log x$ is injective on $x \in \mathbb{R}^p_{>0}$ and $\|\cdot\|_2$ is the standard Euclidean norm.

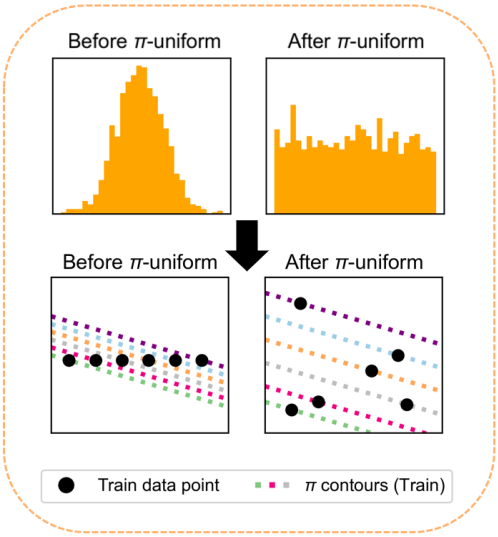

Figure 7: The schematic of π-uniform strategy

Remark 1. *Under the multiplicative scaling action as in Section 3, we have*

$$d(\rho(v,x),\rho(v,y)) = \|(z+v)-(w+v)\|_2 = \|z-w\|_2 = d(x,y),$$

Hence, the metric d is invariant under this group action.

Remark 2. *The metric d induces the standard point-to-set distance from a point $w = \log y$ to the π-equivalence class $[x]_\pi$ (or equivalently $[z]_\pi = \mathcal{E}(z)$) of a fixed $z = \log x$ as*

$$\operatorname{dist}(y,[x]_\pi) = \operatorname{dist}(w,\mathcal{E}(z)) = \inf_{v\in\ker\Phi^\top}\|w-(z+v)\|_2.$$

Then, define the class-to-class (quotient) distance by

$$\operatorname{dist}([y]_\pi,[x]_\pi) = \operatorname{dist}(\mathcal{E}(w),\mathcal{E}(z)) = \inf_{u,v\in\ker\Phi^\top}\|(w+u)-(z+v)\|_2.$$

Since d is invariant under translations by $-u \in \ker(\Phi^\top)$, this reduces to the point-to-class distance:

$$\operatorname{dist}\big(\mathcal{E}(w),\mathcal{E}(z)\big) = \inf_{u,v\in\ker\Phi^\top}\|w-(z+v-u)\|_2 = \inf_{s\in\ker\Phi^\top}\|\,w-(z+s)\,\|_2 = \operatorname{dist}(w,\mathcal{E}(z)).$$

C π-UNIFORM STRATEGY

Figure 7 shows the schematic of π-uniform strategy. The dense π-contours in train sample are distributed uniformly by tuning dominant-scale input parameters (e.g., $\bar{q}$, $\bar{f}$ in thermal and elasticity case, respectively) so that the set of centroids can cover the distribution of all train samples when clustering is performed.

D CLUSTERING PROJECTION

The clustering projection proceeds as follows: (1) Uniformly adjust the pi distribution of the train samples using the π-uniform strategy. (2) Perform K-means clustering based on the π values and determine the centroid for each cluster (marked with a star symbol). (3) Find the closest centroid for a given test sample and project the test sample to the closest position on the π_{test} plane (π-invariant projection).

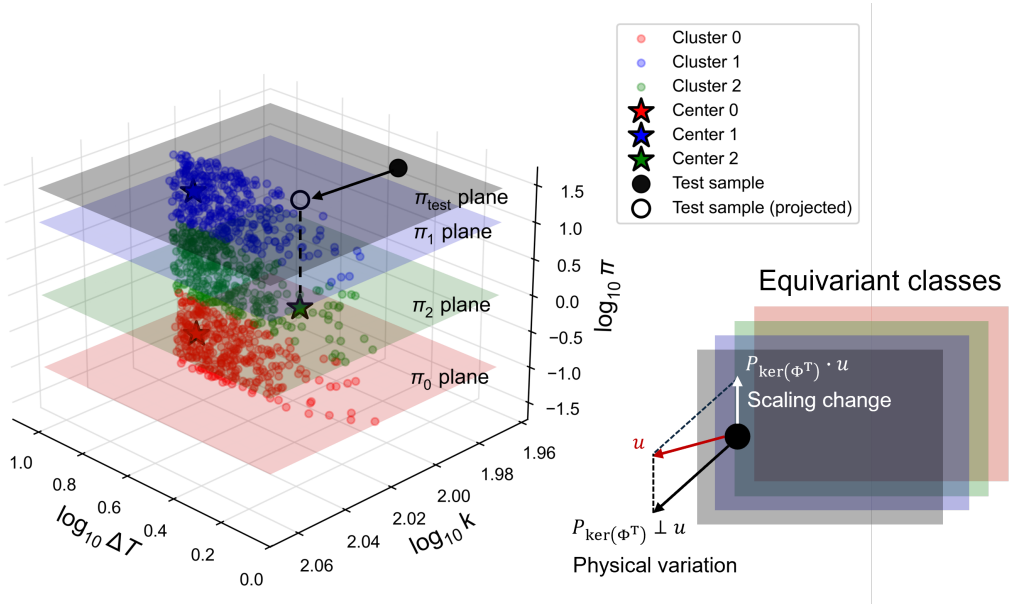

Figure 8: The scheme of clustering projection. Train samples within the same cluster are represented in the same color with centers. The test sample is projected to the nearest center of the cluster while preserving its π_{test}; it moves on the π_{test} plane.

E Experiment metrics

E.1 MAE and RMSE evaluation

In this experiment, image-wise metric is used for evaluating the MAE, RMSE. For the each image matrix $\{Y_{ij}\}$, MAE can be calculated by

$$\mathrm{MAE}_n = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \left| \hat{Y}_{ij}^{(n)} - Y_{ij}^{(n)} \right|. \tag{16}$$

Here, H and W denote the image's height and width, respectively. (in this experiment image correspond to color map). using equation 16, image-wise average MAE can be calculated by

$$\mathrm{MAE} = \frac{1}{N} \sum_{i=1}^{N} \mathrm{MAE}_n \tag{17}$$

Additionally, RMSE can be calculated by

$$\mathrm{RMSE}_n = \sqrt{\frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \left(\hat{Y}_{ij}^{(n)} - Y_{ij}^{(n)} \right)^2}. \tag{18}$$

from equation 17, image-wise average RMSE can be calculated by

$$\mathrm{RMSE} = \frac{1}{N} \sum_{i=1}^{N} \mathrm{RMSE}_n \tag{19}$$

F Additional PDE exmaple : Navier-Stokes equation

We did experiments on Navier-Stokes equation and proved that the projection method also holds for nonlinear problems. The steady governing equations are

$$\rho(\mathbf{u}\cdot\nabla)\mathbf{u} = -\nabla p + \mu\nabla^2\mathbf{u} + \mathbf{f} \quad \text{in } \Omega, \tag{20}$$

$$\nabla\cdot\mathbf{u} = 0 \quad \text{in } \Omega, \tag{21}$$

where $\mathbf{u} = (u_x, u_y)$ is the velocity, p is the pressure, ρ and μ are density and dynamic viscosity, and $\mathbf{f}$ is a body force. We consider a domain Ω with Dirichlet velocity boundary conditions $\mathbf{u} = \mathbf{u}_{\mathrm{bc}}$ on $\partial\Omega_D$. The kinematic viscosity is $\nu = \mu/\rho$, and the Reynolds number is $\mathrm{Re} = \rho\Delta U L/\mu$ where ΔU is the difference between the maximum and minimum value of $U = \sqrt{u_x^2 + u_y^2}$ on boundaries. This Reynolds number (Re) is used for dimensionless number of the π-group. We declare the convergence to a steady state when the quantity, $\|\mathbf{u}_x^{t+1} - \mathbf{u}_x^t\|_2 + \|\mathbf{u}_y^{t+1} - \mathbf{u}_y^t\|_2$ (implemented as the sum of the ℓ_2 norms of u_x and u_y where $\Delta t = 10^{-4}$) drops below a tolerance of 10^{-6}.

F.1 Navier-stokes equation (an ideal case unaffected by external forces)

To represent the ideal case, $\mathbf{f}$ is set to 0, assuming that external forcing such as gravity and flow is not present. In this case, the steady governing equations are

$$\rho(\mathbf{u}\cdot\nabla)\mathbf{u} = -\nabla p + \mu\nabla^2\mathbf{u} \quad \text{in } \Omega, \tag{22}$$

$$\nabla\cdot\mathbf{u} = 0 \quad \text{in } \Omega, \tag{23}$$

For the model training, input data is encoded with 6 channels $\tilde{X}$ =[ρ, μ, $\mathbf{u}_{\mathrm{bc}}^x$, $\mathbf{u}_{\mathrm{bc}}^y$, dx, dy], output data is encoded with 2 channels $[u_x, u_y]$. The scaling coefficients α, β, γ and δ are used to scale the parameters while preserving the Re.

$$\rho' = \alpha\rho, \quad \mu' = \beta\mu, \quad \Delta U' = \gamma\Delta U, \quad L' = \delta L \tag{24}$$

For preserving the Re, scaling coefficients should follow the constraint.

$$\mathrm{Re} = \frac{\rho' L' \Delta U'}{\mu'} = \frac{(\alpha\rho)(\delta L)(\gamma\Delta U)}{(\beta\mu)} = \frac{\rho L \Delta U}{\mu} \tag{25}$$

For efficient calculation, log-scale linearization is applied. After the π-invariant projection, we rescale the test input channel-wise $\tilde{X}^* = \tilde{X} \odot \exp(v^*)$. Then

$$\tilde{X}^* = [\alpha\rho, \beta\mu, \gamma\mathbf{u}_{\mathrm{bc}}^x, \gamma\mathbf{u}_{\mathrm{bc}}^y, \delta dx, \delta dy] \tag{26}$$

By using the ΔU scale factor γ, output $\mathbf{u} = (u_x, u_y)$ can be predicted. We observe that applying the π-invariant projection to the OOD data yields higher performance than making predictions without the projection (see Table 2). *Top-3 best* and *Top-3 worst* OOD cases for Navier-Stokes are shown in Figure 9.

Table 2: Test scores on Navier–Stokes. The best scores are bolded.

Method	**MAE**	**RMSE**	**Time**
CNN	0.063±0.004	0.100±0.007	-
CNN + Pairwise Projection	**0.016±0.001**	**0.022±0.001**	27.97±0.9
CNN + π-uniform + 10-Randoms	0.018±0.005	0.024±0.001	**0.91±0.03**
CNN + π-uniform + 10-Centroids	0.017±0.001	0.023±0.001	0.97±0.02
U-Net	0.11±0.007	0.15±0.01	-
U-Net + Pairwise Projection	**0.02±0.001**	**0.02±0.001**	28.23±0.76
U-Net + π-uniform + 10-Randoms	0.03±0.002	0.037±0.002	**0.92±0.02**
U-Net + π-uniform + 10-Centroids	**0.02±0.001**	0.03±0.001	0.98±0.02
FNO	0.034±0.003	0.044±0.003	-
FNO + Pairwise Projection	**0.013±0.001**	**0.018±0.001**	28.49±0.77
FNO + π-uniform + 10-Randoms	0.016±0.001	0.021±0.001	**1.01±0.04**
FNO + π-uniform + 10-Centroids	0.014±0.001	0.019±0.001	1.07±0.05

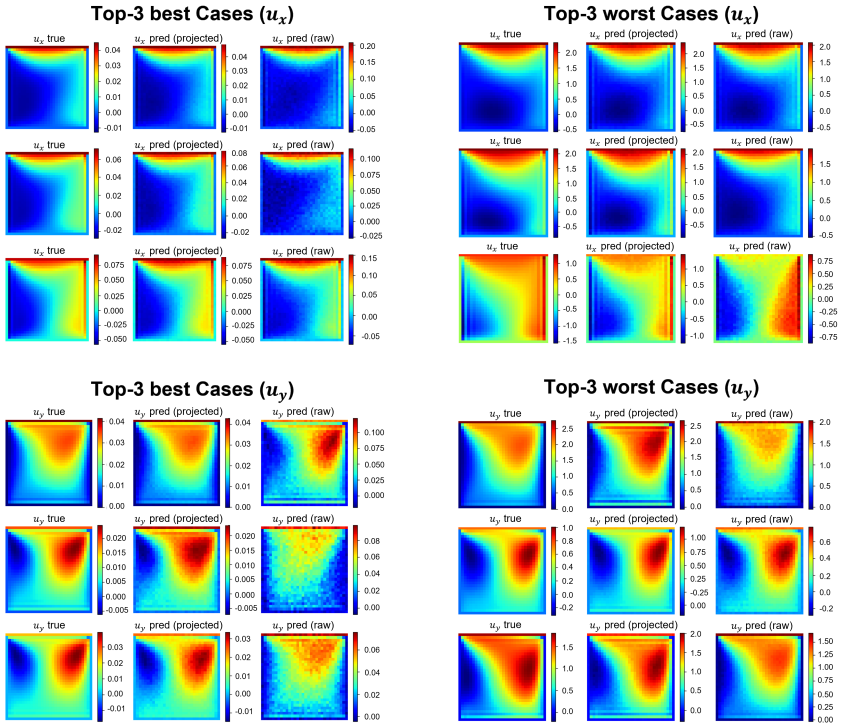

Figure 9: Comparison of ground truths and predictions with *Top-3 best* (a) and *Top-3 worst* (b) cases in Navier-Stokes, under baseline (raw input) vs. Buckingham π-invariant projection.

F.2 Navier-Stokes equation (affected by external forces)

The ideal form of the Navier–Stokes equation does not include gravity or any externally imposed flow ($\mathbf{f} = 0$), additional considerations are required when external forcing is present ($\mathbf{f} \neq 0$). However, the π-groups used in the Buckingham–π projection are derived under this ideal assumption ($\mathbf{f} = 0$). Therefore, when $\mathbf{f} \neq 0$, the resulting π-groups become incomplete.

We also performed experiments to examine how this incompleteness affects the results when such incomplete π-groups are used. In this setup,

$$\mathbf{f} = \frac{8\mu U}{\rho L^2} \tag{27}$$

is used, where μ is the viscosity, U is the boundary-condition velocity, L is the characteristic length, and ρ is the density. The characteristic length L was computed as the product of the grid spacing (Δx) and the resolution.

Table 3: Test Score Comparison for Navier-Stokes ($\mathbf{f} \neq 0$ vs. $\mathbf{f} = 0$) with 10 different test sets. The best scores are bolded.

Method	Navier-stokes ($\mathbf{f} \neq 0$)			Navier-stokes ($\mathbf{f} = 0$)		
	MAE	RMSE	Time	MAE	RMSE	Time
CNN	0.063 ± 0.008	0.085 ± 0.010	-	0.063 ± 0.004	0.100 ± 0.007	-
CNN + Pairwise Projection	$\mathbf{0.029 \pm 0.002}$	$\mathbf{0.037 \pm 0.003}$	28.01 ± 0.92	$\mathbf{0.016 \pm 0.001}$	$\mathbf{0.022 \pm 0.001}$	27.97 ± 0.9
U-Net	0.105 ± 0.003	0.141 ± 0.006	-	0.11 ± 0.007	0.15 ± 0.01	-
U-Net + Pairwise Projection	$\mathbf{0.061 \pm 0.002}$	$\mathbf{0.081 \pm 0.002}$	28.19 ± 0.80	$\mathbf{0.02 \pm 0.001}$	$\mathbf{0.02 \pm 0.001}$	28.23 ± 0.76
FNO	0.056 ± 0.007	0.071 ± 0.008	-	0.034 ± 0.003	0.044 ± 0.003	-
FNO + Pairwise Projection	$\mathbf{0.025 \pm 0.002}$	$\mathbf{0.032 \pm 0.002}$	28.56 ± 0.73	$\mathbf{0.013 \pm 0.001}$	$\mathbf{0.018 \pm 0.001}$	28.49 ± 0.77

Table 3 shows that even in the incomplete case ($\mathbf{f} \neq 0$), applying the Buckingham- π projection improves prediction accuracy. Although the accuracy is slightly lower compared to the ideal case ($\mathbf{f} = 0$), this result demonstrates that the Buckingham-π projection can still enhance prediction performance under incomplete conditions.

G Out-of distribution data in dimensionless space

In this experiment, we evaluate how much the improvement offered by the Buckingham–π invariant projection degrades when the dimensionless numbers extracted from the inputs of OOD data lie outside the π-distribution obtained from the training data, depending on how far they deviate. For the experiment, we use the Reynolds number (Re)—one of the π-groups commonly appearing in the Navier–Stokes equations—as the domain variable. This choice is motivated by the fact that the flow behavior changes significantly with the Reynolds number (Re), making it a suitable quantity for explaining and interpreting the results of our experiment.

For example, when the Re is below 1, the flow is dominated by viscous effects and typically exhibits Stokes flow. When the Re is in the range of 10–40, attached vortices commonly appear, while for Re greater than 40, the flow is dominated by the Von Kármán vortex street. In our setup, the model is trained only on data with Re in the range 0.04-1.01 (Train data), and tested on three datasets: Test-ID (0.04–1.02), Test-OOD (mid) with a range of 0.32–27.06, and Test-OOD (extreme) with a range of 42.28–63.43 (shown in Figure 10).

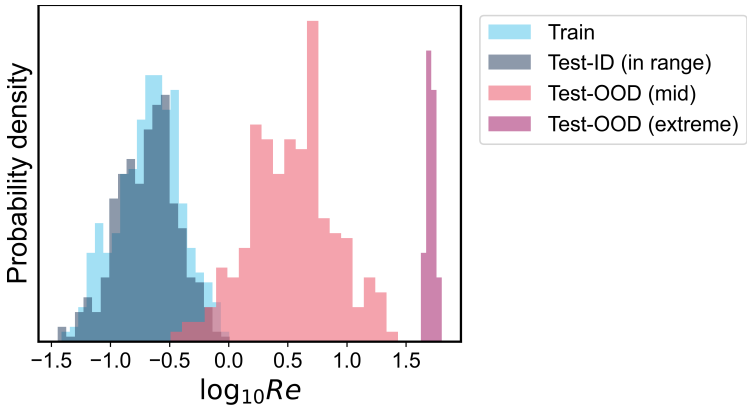

Figure 10: The distribution of $\log_{10}\mathrm{Re}$. Train Re range: 0.04-1.01, Test-ID Re range: 0.04-1.02, Test-OOD (mid) Re range: 0.32-27.06, Test-ID (extreme) Re range: 42.28-64.43.

Table 4: Test scores on Navier-Stokes with 3 different test sets. The best scores are bolded.

Method	**Test-ID (in-range)**			**Test-OOD (mid)**			**Test-OOD (extreme)**		
	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE	R2
CNN	0.02	0.03	0.02	0.20	0.28	0.12	0.38	0.59	-0.04
CNN + Pairwise Projection	**0.005**	**0.009**	**0.82**	**0.08**	**0.13**	**0.76**	**0.34**	**0.52**	**0.18**
U-Net	0.01	0.02	0.50	0.23	0.30	0.10	0.38	0.54	0.18
U-Net + Pairwise Projection	**0.003**	**0.004**	**0.97**	**0.06**	**0.09**	**0.92**	**0.24**	**0.32**	**0.70**
FNO	0.006	0.007	0.95	0.20	0.26	0.39	0.33	0.45	0.14
FNO + Pairwise Projection	**0.002**	**0.002**	**0.99**	**0.04**	**0.05**	**0.97**	**0.21**	**0.27**	**0.78**

Table 4 shows that as the test dataset moves farther away from the training dataset, the prediction accuracy gradually decreases despite the improvements provided by the π-invariant projection. For the CNN model, the Test-ID dataset—which is similar to the training distribution—achieves an R^2 score of 82%. However, the score drops to 76% for the Test-OOD (mid) dataset, and further down to 18% for the Test-OOD (extreme) dataset. This trend appears consistently across different models, including U-Net and FNO. Nevertheless, an interesting observation from these results is

that even in cases where the data lie far outside the π-distribution, using the π-invariant projection still improves prediction performance compared to making predictions without it. Although the degree of improvement diminishes when the test dataset deviates too far from the training data, the π-uniform strategy proposed in the paper provides a way to uniformly cover this range. Moreover, in practical settings, the physically meaningful π-range can be predefined based on prior knowledge from physics and engineering.

H ALGORITHM

Algorithm 1 Buckingham-π-invariant test-time projection

Train surrogate f_θ on $\mathcal{D}_{\text{train}} = \{(X_i, Y_i)\}_{i=1}^{M}$
for each test pair $(\tilde{X}_i, \tilde{Y}_i) \in \mathcal{D}_{\text{test}}$ **do** ▷ OOD test set, $\mathcal{D}_{\text{test}} = \{(\tilde{X}_i, \tilde{Y}_i)\}_{i=1}^{M}$
    $\tilde{z}_i \leftarrow \log \psi(\tilde{X})$ ▷ ψ=feature extractor
    Set $v_i^t = z_i - \tilde{z}$ and decompose v_i^t into $P_\parallel v_i^t$ (intra-class) and $(I - P_\parallel)v_i^t$ (inter-class).
    Select the nearest train sample by
    $\arg\min_{i=1,\ldots,M} \|v_i^t - v_i^*\|_2$ where $v_i^* = P_\parallel v_i^t$
    Form projected input $\tilde{X}^* = \tilde{X} \odot \exp(v_{i^*}^*)$; and predict $\tilde{Y}_{\text{scaled}} \leftarrow f_\theta(\tilde{X}^*)$
    $\tilde{Y}_{\text{orig}} \leftarrow \textsc{InverseScale}(\tilde{Y}_{\text{scaled}}, v_{i^*}^*)$ ▷ e.g., $T = T_{\min} + \gamma^{-1}(T_{\text{scaled}} - T_{\min})$;
end for