UNIMANO: NEURAL OPERATOR-BASED UNIFIED MATERIAL REPRESENTATIONS FOR GENERALIZABLE THERMAL MODELING

Anonymous authors
Paper under double-blind review

ABSTRACT

Identifying material thermophysical properties from limited temperature observations and predicting subsequent thermal evolution are key goals of thermal modeling. Neural networks offer new approaches, but methods based on per-instance reconstruction still require independent optimization for each material and lack a unified representation that can be shared and transferred across materials. To address this limitation, we propose UniMaNO: Neural Operator-Based Unified Material Representations for Generalizable Thermal Modeling, a neural operator framework centered on a unified material representation. Specifically, UniMaNO maps temperature observations to material thermophysical property functions through a two-stage framework: first constructing a shared thermophysical function basis from training conductivity curves, then using temperature reconstruction errors from a differentiable heat conduction simulator to learn the observation-to-material representation mapping and refine material representations. After training, UniMaNO identifies the thermophysical properties of unseen materials without per-instance optimization, reuses the inferred representations for thermal evolution prediction under new geometries and heat-source conditions, and further supports few-shot refinement using the input observations. Extensive experiments demonstrate that UniMaNO enables rapid property identification and physically consistent thermal evolution prediction across materials and operating conditions.

1 INTRODUCTION

Thermal modeling is fundamental to material characterization (Parker et al., 1961), thermal system monitoring, and predictive simulation (Zálešák et al., 2023; Zobeiry & Humfeld, 2021). Yet identifying material thermophysical properties that support accurate thermal evolution prediction remains challenging when only limited temperature observations are available (Jo et al., 2024; Shang et al., 2025). Although neural networks (Karniadakis et al., 2021) have advanced both thermophysical property identification (Qiu et al., 2025; Moon et al., 2025) and temperature-field prediction (Cai et al., 2021; Ding et al., 2026), optimizing these tasks separately does not directly constrain the thermal evolution produced by the inferred properties. These properties may therefore fail to reproduce the material's thermal behavior in subsequent simulations.

Per-instance reconstruction methods connect thermophysical property identification with thermal evolution by fitting observed temperature responses through physical constraints (Raissi et al., 2019; Karnakov et al., 2023) or differentiable simulation. The recovered properties are thus constrained by the forward thermal evolution they produce. However, independently modeling and optimizing each instance does not fully exploit shared structure across the thermophysical property functions of different materials. Identifying each new material still requires solving a separate inverse problem, limiting knowledge transfer and efficient generalization across materials.

To address this limitation, we propose **UniMaNO**, a neural operator learning framework (Wang et al., 2021; Li et al., 2024) for thermal modeling that represents diverse materials within a unified material representation space. UniMaNO directly identifies the thermophysical properties of unseen materials from limited temperature observations and predicts thermal evolution under different geometries and heat-source conditions. Inspired by the universal approximation theorem for operators (Chen & Chen,

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

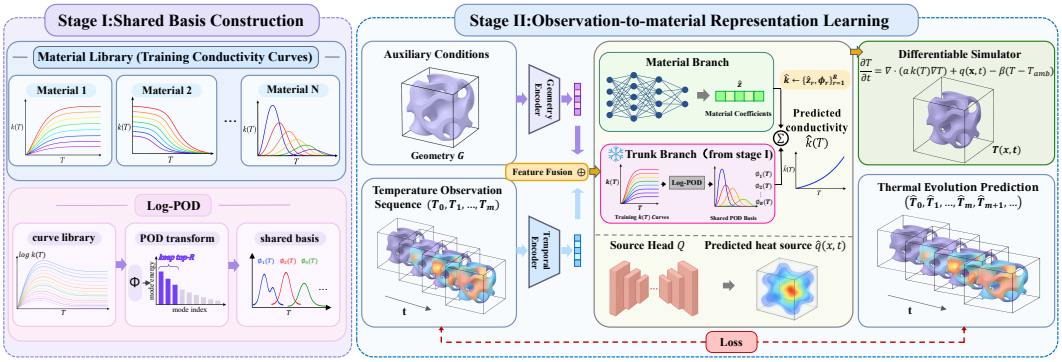

Figure 1: **Overview of UniMaNO.** Stage I constructs a shared thermophysical function basis from training data. Stage II learns to map temperature observations to material representations through temperature sequence reconstruction.

1995; Lu et al., 2021; Lanthaler et al., 2022), we formulate thermophysical property identification as learning a mapping between function spaces: from temperature observations, conditioned on the physical setting, to material thermophysical property functions. The unified material representation space captures shared structure across these functions while preserving material-specific differences. This allows diverse materials to be identified within one framework and their inferred thermophysical property functions to be reused under new physical conditions.

We learn this mapping through a two-stage framework (Kontolati et al., 2024). The first stage constructs shared structure from thermophysical property functions across materials to establish the unified material representation space (Lu et al., 2022). The second learns the mapping from temperature observations to material representations through feedback from a differentiable heat conduction simulator. Predicted material representations are decoded into thermophysical property functions and supplied to the simulator together with the physical conditions. Discrepancies between simulated and observed temperatures are backpropagated through the simulator to train the identification model. Material representations are therefore constrained by both the shared thermophysical structure and the thermal evolution they produce. After training, UniMaNO identifies the thermophysical properties of unseen materials without per-instance optimization and reuses the inferred representations for zero-shot thermal evolution prediction under new geometries and heat-source conditions. The identification results can also be refined through few-shot adaptation using a few temperature frames already included in the input observations (Goswami et al., 2022). Our main contributions are:

- We propose UniMaNO, a neural operator framework that combines unified material representations with differentiable simulation feedback for direct thermophysical property identification and cross-condition thermal evolution prediction.
- We construct MatTherm: A Benchmark for Material Identification and Thermal Evolution Prediction, spanning diverse temperature-dependent conductivity functions, geometries, and heat sources in 2D and 3D, with controlled generalization protocols.
- Experiments demonstrate that UniMaNO efficiently identifies thermal conductivity functions, improves the accuracy and physical consistency of thermal evolution prediction, and achieves zero-shot generalization to unseen materials, geometries, and heat-source conditions.

2 Related Work

2.1 Physics-Informed Thermal Modeling and Neural Operators

Physics-informed neural networks incorporate governing equations and initial and boundary conditions into thermal modeling (Cai et al., 2021; 2020; He et al., 2021), while neural operators (Kovachki et al., 2023) such as DeepONet (Lu et al., 2021), FNO (Li et al., 2021; Kovachki et al., 2021), and PINO (Li et al., 2024) learn response mappings across varying input conditions (Li et al., 2023). In forward heat conduction modeling (Koric & Abueidda, 2023), these methods typically take material

properties and external conditions as inputs to predict temperature evolution, enabling efficient prediction of parameterized thermal processes (Zobeiry & Humfeld, 2021). Our work addresses the complementary mapping from thermal observations to material functions and uses the inferred functions for subsequent thermal prediction.

2.2 Inverse Thermal Property Identification

Inverse thermal property identification recovers unknown material properties from temperature observations (Asadzadeh et al., 2023; Tang et al., 2025; Bidou et al., 2026) through approaches such as inverse PINNs (Raissi et al., 2019; Jo et al., 2024; Yang et al., 2021; Mishra & Molinaro, 2022), PDE-constrained optimization (Karnakov et al., 2023; 2024), and learning-based parameter estimation (Qiu et al., 2025; Zhang & Li, 2025). Instance-specific methods typically require optimization for each new observation (Oommen & Srinivasan, 2022; Jiang et al., 2023), whereas recent inverse neural operators support inference of temperature-dependent thermal properties for unseen materials (Moon et al., 2025). Building on this direction, we first construct a shared representation of thermal conductivity functions across materials, then learn a mapping from transient thermal observations to this representation. The inferred material functions can be directly integrated into a physical solver and reused under different operating conditions.

2.3 Differentiable Physics-Based Inversion

Differentiable physics-based inversion constrains material identification by minimizing discrepancies between simulated and observed responses. For example, NeFTY (Zhong et al., 2026) represents a thermal diffusivity field as a neural field and optimizes it through a differentiable heat-transfer solver. However, this instance-specific approach fits a separate material field for each scene without jointly learning a shared observation-to-material mapping across material instances. In contrast, we jointly train a unified material operator across multiple materials, representing their constitutive behavior within a shared function space. By retaining physical simulation feedback during training (Hu et al., 2020; Um et al., 2020), our approach enables direct material identification from observations and reuse across operating conditions, without separate optimization for each new material.

3 Methodology

3.1 Problem Formulation and Overview

We aim to learn unified material representations that support thermophysical property identification and thermal evolution prediction for unseen materials. Consider a temperature field $T(\mathbf{x}, t)$ in a domain Ω. Its thermal evolution is governed by

$$\frac{\partial T}{\partial t} = \nabla \cdot \big(a\, k(T) \nabla T\big) + q(\mathbf{x}, t) - \beta\big(T - T_{\mathrm{amb}}\big). \tag{1}$$

Here, $k(T)$ is the material's thermal conductivity function, and $q(\mathbf{x}, t)$ is an effective heat source expressed as a temperature rate. The known coefficient a converts conductivity to diffusivity; β and T_{amb} are the known ambient cooling rate and ambient temperature, respectively.

Let $u = (T_0, T_1, \ldots, T_m) \in \mathcal{U}$ denote the temperature observation sequence used for identification, where T_i is the temperature field at frame i. Given the geometry and boundary conditions, we formulate thermal conductivity identification as an operator mapping from the temperature observation space to the thermophysical property function space. Identification uses material and heating priors learned during training, without test-time heat-source measurements. Omitting the known conditions, we write:

$$\mathcal{G}_\theta : \mathcal{U} \to \mathcal{K}, \qquad \mathcal{G}_\theta(u) = \widehat{k}. \tag{2}$$

Here, $\mathcal{K}$ consists of the thermal conductivity functions considered in this work. To account for the effect of unknown heating conditions on the temperature response, an auxiliary heat source branch estimates $\widehat{q}$. This estimate conditions conductivity identification and supplies the heat source for temperature reconstruction and future thermal evolution prediction under the original operating conditions. For prediction under new operating conditions, the inferred conductivity function is paired with the prescribed heat source of the new setting, as illustrated in Figure 1.

We first describe the parameterization of the thermophysical property operator (Section 3.1). We then combine function supervision with forward simulation feedback through two stages of learning (Section 3.2). Finally, we describe zero-shot thermophysical property identification and few-shot adaptation, together with their use for thermal evolution prediction (Section 3.3).

Thermophysical property operator parameterization. Inspired by operator approximation theory, we represent log conductivity using shared basis functions and material-specific coefficients, adopting the branch–trunk formulation of DeepONet (Lu et al., 2021) to parameterize $\mathcal{G}_\theta$ as a mapping from limited temperature observations to conductivity functions.

Given M training samples $\{k_i\}_{i=1}^M$, we construct R shared basis functions $\{\phi_r\}_{r=1}^R$ from their log representations, which the trunk evaluates at a query temperature ξ. The branch combines spatiotemporal features of temperature observations u with the estimated heat source $\widehat{q}$ to predict material coefficients $\widehat{z} \in \mathbb{R}^R$, yielding:

$$k(\xi) \approx \mathcal{G}_\theta(u)(\xi) = \exp\left[\phi_0(\xi) + \sum_{r=1}^{R} \widehat{z}_r \, \phi_r(\xi)\right]. \tag{3}$$

Here, ϕ_0 is a fixed offset function, and the exponential ensures positive conductivity. The coefficients $\widehat{z}$ form the material representation in the unified space $\mathcal{Z}$, while the output function $\mathcal{G}_\theta(u)$ belongs to the thermophysical property function space $\mathcal{K}$.

Trunk: shared thermophysical property basis functions. The trunk captures shared structure of thermophysical properties across materials so that different conductivity functions can be expressed using the same basis. We adopt the fixed-basis formulation of POD-DeepONet (Lu et al., 2022), which uses proper orthogonal decomposition (POD) modes extracted from training functions as the trunk. We apply this formulation in the log-conductivity domain to obtain a log-POD representation. The trunk evaluates the resulting basis functions at each query temperature. These functions are constructed from the known conductivity curves in Stage I and remain fixed in Stage II; the branch coefficients account for differences between materials.

Branch: heat-source-conditioned material representations. The branch predicts the coefficients of a material's conductivity function from limited temperature observations. To avoid attributing temperature-response differences caused by heating conditions to material properties, we condition coefficient prediction on an estimated heat source. Specifically, the temporal encoder E_{temp} encodes the temperature sequence u into spatiotemporal features H, from which the auxiliary branch Q predicts a nonnegative heat source field $\widehat{q} = Q(H)$. We fuse the estimated source with H and apply spatial pooling to obtain heat-source-conditioned frame features v_t. Temporal attention then aggregates these features across the observation window:

$$\overline{v} = \sum_t \frac{\exp(s_t)}{\sum_j \exp(s_j)} \, v_t, \tag{4}$$

where s_t is the attention score for frame t. Finally, the geometry encoder E_{geo} encodes geometry g into features $e_g = E_{\mathrm{geo}}(g)$. These features are fused with the aggregated features $\overline{v}$ and passed through a multilayer perceptron to obtain the material representation $\widehat{z}$.

3.2 Learning with Function Supervision and Simulation Feedback

Learning the thermophysical property operator directly through temperature sequence reconstruction is challenging. To reduce the learning difficulty, we adopt a two-stage strategy. We first construct a shared function basis from known conductivity curves. We then fix this basis and learn the observation-to-material representation mapping through supervision from reference material representations and temperature reconstruction feedback, while jointly training the heat source branch.

Stage I: shared thermophysical function basis construction. This stage extracts shared structure from the conductivity curves of training materials and constructs a set of basis functions shared across materials. Given training curves $\{k_i\}_{i=1}^M$, we evaluate their logarithms on a common temperature

grid to obtain $\boldsymbol{\ell}_i$. Applying POD to the centered curves yields the leading R basis vectors $\{\boldsymbol{v}_r\}_{r=1}^{R}$. Standardizing the projection coefficients gives the reference material representation:

$$z_{ir}^{\star} = \frac{\langle \boldsymbol{\ell}_i - \boldsymbol{\mu}, \boldsymbol{v}_r \rangle - m_r}{s_r}, \qquad r = 1, \ldots, R. \tag{5}$$

Here, $\boldsymbol{\mu}$ is the mean log-conductivity vector, and m_r and s_r are the mean and standard deviation of the r-th projection coefficient across training curves.

The corresponding shift and scaling are absorbed into the shared trunk, so $z_i^{\star}$ directly provides the expansion coefficients in Section 3.1. Stage II fixes this trunk and uses $z_i^{\star}$ as representation supervision, together with forward simulation feedback, to learn the mapping from temperature observations to material representations.

Differentiable heat conduction simulation and observation feedback. To relate the predicted conductivity function and heat source to temperature observations, we use a differentiable heat conduction simulator to reconstruct the observations. For scene i, the model predicts the conductivity function $\widehat{k}_i = \mathcal{G}_\theta(u_i)$ and heat source $\widehat{q}_i$ from observations u_i. Given geometry g_i and boundary conditions, the simulator $\mathcal{S}_{g_i}$ starts from the initial temperature field $T_0^{(i)}$, with the reconstructed temperatures and reconstruction loss defined as:

$$\widehat{\mathbf{T}}^{(i)} = \mathcal{S}_{g_i}\left(T_0^{(i)}; \mathcal{G}_\theta(u_i), \widehat{q}_i\right), \qquad \mathcal{L}_{\text{sim}}^{(i)} = \ell_{\text{obs}}\left(\widehat{\mathbf{T}}^{(i)}, \mathbf{T}_{\text{obs}}^{(i)}\right). \tag{6}$$

Here, ℓ_{obs} evaluates reconstruction error at valid observation locations. At each time step, the simulator evaluates conductivity at the current temperature and advances thermal evolution using the resulting conductivity, predicted heat source, and known ambient cooling term. The reconstruction error is backpropagated through the simulator to update the material representation and heat source prediction networks.

Stage II: observation-to-material representation learning. Using the procedure above, we learn unified material representations through temperature reconstruction error. The shared basis functions and reference material representations $z_i^{\star}$ obtained in Stage I remain fixed, while all other network components are trained. The predicted conductivity function and heat source are used to reconstruct temperature observations. The reconstruction error is backpropagated through the simulator to update the material representations. In addition to reconstruction feedback, we supervise the predictions using ground-truth heat sources and reference material representations, respectively. We also introduce a consistency constraint to encourage consistent representations of the same material across conditions. The joint training objective is:

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} \Big[\lambda_{\text{sim}} \mathcal{L}_{\text{sim}}^{(i)} + \lambda_q \mathcal{L}_q(\widehat{q}_i, q_i^{\star}) + \lambda_z^{\text{II}} \mathcal{L}_z(\widehat{z}_i, z_i^{\star}) \Big] + \lambda_{\text{inv}} \mathcal{L}_{\text{inv}}\big(\{\widehat{z}_i\}_{i=1}^{N}\big). \tag{7}$$

Here, θ denotes all trainable network parameters, and N is the number of training scenes. For scene i, $q_i^{\star}$ is the ground-truth heat source, and $z_i^{\star}$ is the reference representation of its material obtained in Stage I. The losses $\mathcal{L}_q$ and $\mathcal{L}_z$ measure heat source and material representation prediction errors, respectively. The consistency loss $\mathcal{L}_{\text{inv}}$ measures representation differences within groups of scenes sharing the same material. The λ coefficients weight the corresponding loss terms.

Temperature reconstruction and future thermal evolution prediction under the original operating conditions both use the model-predicted heat source $\widehat{q}$. For prediction under new operating conditions, the inferred conductivity function is combined with the prescribed heat source of the new setting.

3.3 Inference and Few-Shot Adaptation

Zero-shot thermophysical property identification. Given limited temperature observations u of a target scene, the model predicts a material representation in a single forward pass. The operator outputs the conductivity function $\widehat{k} = \mathcal{G}_\theta(u)$. This procedure requires no per-instance optimization.

Few-shot adaptation. We refine each scene's identification using observations already present in u. The complete input u is retained for identification, while selected frames from u, indexed by $\mathcal{I}$,

supervise adaptation. We initialize the network parameters θ with the pretrained parameters θ_0 and keep the shared basis functions and simulator fixed while adapting the remaining network parameters. Using the current conductivity function and heat source predictions, we reconstruct temperature observations and update the network by minimizing:

$$\min_{\theta} \ \ell^{\mathcal{I}}_{\mathrm{obs}}\Big(\widehat{\mathbf{T}}_{\theta}, \mathbf{T}_{\mathrm{obs}}\Big) + \frac{\lambda_{\mathrm{prior}}}{n_{\theta}} \left\|\theta - \theta_0\right\|_2^2 . \tag{8}$$

Here, $\ell^{\mathcal{I}}_{\mathrm{obs}}$ evaluates reconstruction error at valid observation locations in the selected frames, and n_θ is the number of optimized scalar parameters. The second term limits parameter drift from θ_0. Each scene is adapted independently from θ_0. If adaptation yields a higher reconstruction error on the selected frames than zero-shot inference, we retain the zero-shot identification result.

The conductivity inferred by zero-shot identification or few-shot adaptation is used for thermal prediction. Under the original conditions, simulation starts from the temperature field at the beginning of the prediction interval and uses the model-predicted heat source. For inference under new operating conditions, we reuse the inferred conductivity function with the heat source provided for that setting, together with its geometry and initial temperature field.

4 Experiments

4.1 Experimental Setup

Data and protocols. We construct MatTherm, a 2D and 3D benchmark for material identification and thermal evolution prediction across diverse materials, geometries, and heat-source conditions. The synthetic material library contains thousands of temperature-dependent conductivity functions, covering five response archetypes and 15 functional subfamilies. Each spatial dimension includes 13 geometry families and 15 heat-source families. Geometries encompass regular, curved, non-convex, and perforated domains. See Appendix A for details.

Baselines. For thermal evolution prediction, we compare with F-FNO (Tran et al., 2023), DPOT (Hao et al., 2024), Poseidon (Herde et al., 2024), MemNO (Buitrago Ruiz et al., 2025), and Transolver3 (Zhou et al., 2026). For material identification, we compare accuracy and efficiency with the per-instance inversion methods Inverse PINN (Raissi et al., 2019), gPINN (Yu et al., 2022), and mODIL (Karnakov et al., 2023), where mODIL performs numerical inverse optimization. Baseline configurations, training and optimization settings are provided in Appendix C.

Metrics. RMSE and response-normalized relative L_2 error measure absolute and relative temperature prediction accuracy, while PDE residuals and energy balance errors assess physical consistency. We evaluate conductivity recovery using relative L_2 error over the full (Full) and observed (Obs.) temperature ranges, and report per-instance identification time to assess efficiency. Complete metric definitions and evaluation procedures are provided in Appendix C.

4.2 Long-Horizon Thermal Prediction in 2D and 3D

Table 1: **Comparison of long-horizon thermal evolution prediction in 2D and 3D.** Results are averaged over 500 seen materials per dimension; lower values are better for all metrics.

	2D				3D			
Method	RMSE	L2	PDE	Energy	RMSE	L2	PDE	Energy
F-FNO	3.82e−01	1.81e−02	1.17e+00	1.39e−01	4.06e−01	2.97e−02	1.77e+00	2.87e−01
DPOT	8.45e−01	3.75e−02	1.50e+00	8.59e−01	1.88e+00	1.45e−01	1.75e+00	1.36e+00
Poseidon	7.00e−01	3.36e−02	1.29e+00	3.75e−01	5.82e+00	4.47e−01	1.85e+00	1.97e+00
MemNO	4.00e−01	1.89e−02	1.18e+00	2.44e−01	3.80e−01	2.85e−02	1.77e+00	2.40e−01
Transolver3	8.88e+00	4.41e−01	1.86e+00	1.96e+00	4.90e+00	3.79e−01	1.84e+00	1.96e+00
UniMaNO	**1.12e−01**	**5.04e−03**	**5.70e−02**	**2.99e−05**	**2.61e−01**	**1.72e−02**	**8.78e−02**	**8.66e−05**

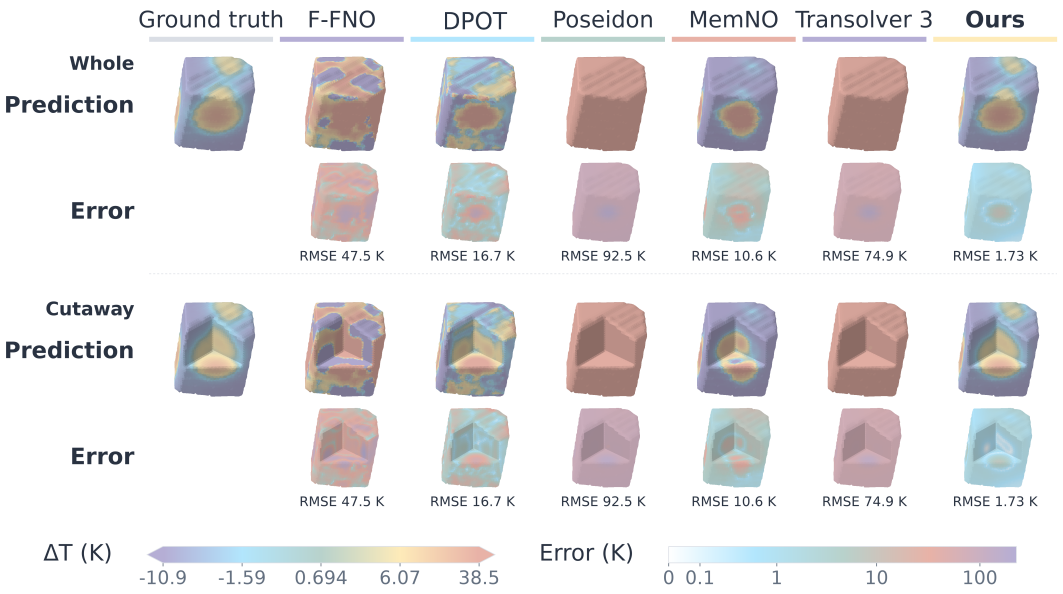

Figure 2: **3D thermal prediction at the final time step.** Whole-volume and cutaway views show predictions and absolute errors, with RMSEs computed for this frame.

Table 2: **Comparison of cross-condition thermal evolution prediction for unseen materials in 2D.** Results are averaged over 500 unseen materials per setting; lower values are better for all metrics.

	Source change		Geometry change		Joint change	
Method	RMSE (K)	L2	RMSE (K)	L2	RMSE (K)	L2
F-FNO	9.50e–01	3.96e–02	7.74e–01	6.00e–02	1.13e+00	5.71e–02
DPOT	3.42e+00	1.35e–01	4.41e+13	2.59e+12	5.83e+00	2.91e–01
Poseidon	3.95e+00	1.55e–01	2.89e+00	1.79e–01	5.05e+00	2.43e–01
MemNO	9.49e–01	3.74e–02	1.56e+00	9.95e–02	1.59e+00	7.99e–02
Transolver3	2.04e+01	8.12e–01	1.36e+01	1.04e+00	1.83e+01	8.73e–01
UniMaNO-ZS	5.63e–02	2.33e–03	5.13e–02	3.44e–03	6.82e–02	3.57e–03
UniMaNO-FS	**3.90e–02**	**1.61e–03**	**3.39e–02**	**2.25e–03**	**4.64e–02**	**2.41e–03**

We train separately on the 2D and 3D datasets. At test time, conductivity inferred from observations of a seen material drives simulation beyond the observation window, without per-instance optimization. UniMaNO achieves the best results on all metrics in both dimensions (Table 1), reducing RMSE by approximately 71% and 31% relative to the strongest baselines. PDE residuals are more than an order of magnitude lower, and energy balance errors remain below 10^{-4}. The shared representation thus supports accurate long-horizon prediction across diverse materials.

4.3 Generalization across Materials and Conditions

We evaluate unseen materials under new operating conditions (Table 2). Although neural operator baselines support cross-condition prediction, their accuracy is sensitive to condition changes: MemNO's RMSE increases from 0.949 K under a heat-source change to 1.59 K under joint geometry and heat-source changes. UniMaNO maintains RMSEs below 0.07 K across all three settings, improving over the strongest baselines by 93.4%–94.1%. Its RMSE differs by less than 0.004 K between seen and unseen materials in each setting, indicating that the shared material space accommodates new materials without instance-specific optimization.

Figure 3 shows that conductivity inferred from one observed sequence can be directly reused when the geometry and heat-source conditions change, with UniMaNO maintaining low prediction errors across different settings. Figure 5 further considers a case where the material, geometry, and heat

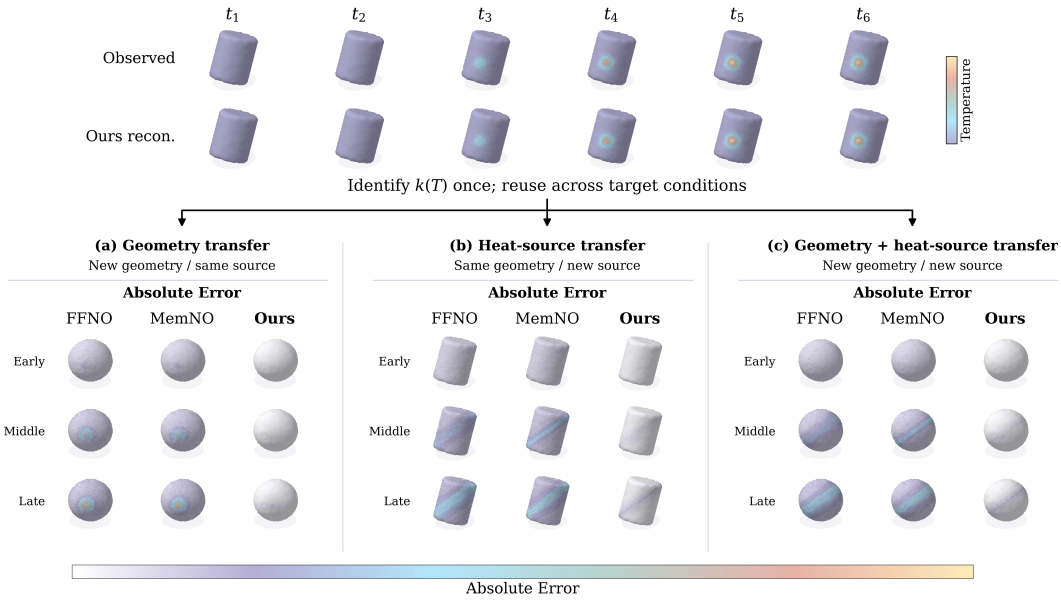

Figure 3: **Cross-condition thermal evolution prediction from a single material identification.** Top: observed and reconstructed temperatures. Bottom: absolute prediction errors when reusing the same inferred $k(T)$ under (a) new geometry, (b) new heat source, and (c) both.

source are all unseen; compared with random or mean material representations, the representation inferred by UniMaNO substantially reduces reconstruction error and accurately recovers the thermal response in this unseen case. These results demonstrate the cross-condition reusability of the inferred conductivity and the effectiveness of material identification in unseen scenarios.

Few-shot adaptation. To assess whether few-shot adaptation can further improve prediction, we adapt the pretrained model through simulation-based reconstruction using a few temperature frames from the pre-prediction observation window. Evaluation uses the same input observations and prediction targets as zero-shot inference. Adaptation budgets and optimization settings are provided in Appendix D.

Table 3: **Comparison of conductivity identification accuracy and efficiency on seen and unseen materials.** Results are averaged over three materials per split. † PDE residuals are evaluated on transfer trajectories using the reference conductivity.

	Seen materials ($n = 3$)				**Unseen materials** ($n = 3$)			
Method	Full L2	Obs. L2	PDE†	Time (s)	Full L2	Obs. L2	PDE†	Time (s)
Inverse PINN	9.26e–01	9.13e–01	1.37e–01	2.01e+02	3.63e+00	8.75e–01	1.11e–01	2.04e+02
gPINN	9.37e–01	9.02e–01	1.31e–01	2.82e+02	3.62e+00	8.86e–01	1.11e–01	2.81e+02
mODIL	8.67e–01	8.67e–01	9.42e–02	1.45e+02	8.18e–01	6.43e–01	3.25e–02	1.44e+02
UniMaNO-ZS	**3.33e–01**	6.21e–02	3.73e–03	**2.48e–01**	2.86e–01	**5.28e–02**	1.06e–03	**1.37e–02**
UniMaNO-FS	3.38e–01	**5.06e–02**	**1.72e–03**	1.91e+00	**2.82e–01**	6.21e–02	**3.58e–04**	1.63e+00

4.4 Conductivity Identification and Material-Space Structure

The learned observation-to-material mapping enables UniMaNO-ZS to improve both conductivity recovery and identification efficiency across seen and unseen materials (Table 3). On unseen materials, it reduces Full/Obs. relative L_2 errors by approximately 65%/92% over mODIL, while accelerating identification by roughly four orders of magnitude compared with per-instance inversion.

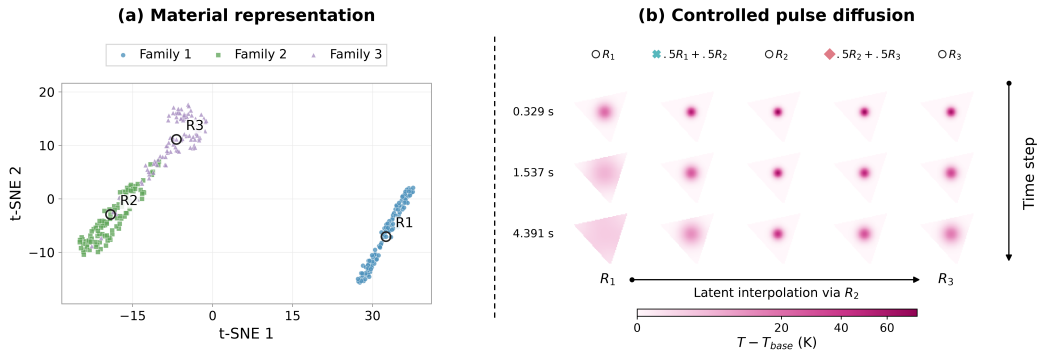

Figure 4: **Material representation analysis.** (a) t-SNE of three material families. (b) Thermal evolution under identical conditions for the marked representations in (a) (columns 1, 3, 5) and midpoint interpolations between adjacent representations (columns 2, 4).

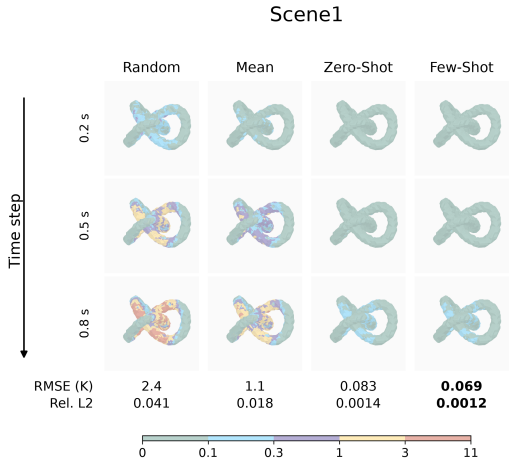

Figure 5: **Material identification and thermal reconstruction for an unseen case.**

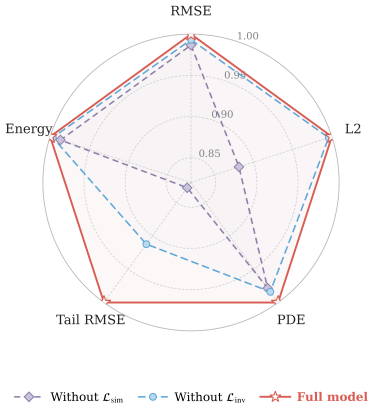

Figure 6: **Comparison of ablation variants.** Error ratio (full/variant); higher is better.

Figure 4 reveals physically meaningful structure in the learned material space. Representations exhibit family-wise grouping in the t-SNE projection (a). Under identical conditions, interpolating between the representations selected in (a) produces heat diffusion patterns between those of the original materials (b).

4.5 Ablation Studies

We remove $\mathcal{L}_{\text{sim}}$ and $\mathcal{L}_{\text{inv}}$ and evaluate all variants zero-shot (Figure 6). Tail RMSE increases by 20.8% without simulation feedback and 9.6% without same-material consistency, compared with RMSE increases of 1.4% and 0.7% under the seen-material heat-source-change setting. These results support both constraints for learning representations that reliably drive thermal evolution.

5 Conclusion

We present UniMaNO, a neural operator framework unifying material identification and thermal prediction through shared material representations. Its two-stage training combines shared thermophysical structure with differentiable simulation feedback to infer conductivity functions from limited temperature observations without per-instance optimization and reuse them across geometries and heat sources. Experiments demonstrate accurate, physically consistent prediction in 2D and 3D and zero-shot generalization to unseen materials and operating conditions.

AI USE STATEMENT

In this work, we used generative AI tools for the following tasks with required disclosure: providing feedback on research methodology and experimental design, assisting with method implementation, assisting with translation, and interpreting experimental results.

We did not use generative AI tools to generate synthetic datasets, develop theoretical models or conceptual frameworks, formulate mathematical claims, provide critical ingredients for proving mathematical claims, assist in writing proofs, propose or refine hypotheses, or clean and reformat datasets. Qualitative and thematic data analysis is not applicable to this work.

Additionally, we used generative AI tools for tasks with recommended disclosure, including editing and drafting parts of the manuscript, searching for and summarizing relevant literature, creating or modifying scientific figures, creating, editing, and checking software code, and improving the structure, readability, and presentation of the manuscript.

We have reviewed all AI-assisted work. AI-assisted method implementations and code were checked and tested by the authors, while AI-assisted translations, experimental suggestions, result interpretations, and manuscript content were verified against the actual methods, experimental results, and relevant source materials. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

REFERENCES

Mohammad Zhian Asadzadeh, Klaus Roppert, and Peter Raninger. Material Data Identification in an Induction Hardening Test Rig with Physics-Informed Neural Networks. *Materials*, 16(14):5013, 2023. doi: 10.3390/ma16145013.

M.S. Bidou, J.-G. Bauzin, and N. Laraqi. Physics-informed neural networks for 2D-transient inverse heat conduction problem involving static and moving heat source. *International Journal of Heat and Mass Transfer*, 259:128349, 2026. doi: 10.1016/j.ijheatmasstransfer.2026.128349.

Ricardo Buitrago Ruiz, Tanya Marwah, Albert Gu, and Andrej Risteski. On the Benefits of Memory for Modeling Time-Dependent PDEs. In *International Conference on Learning Representations*, volume 2025, pp. 54972–55002, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/89379d5fc6eb34ff98488202fb52b9d0-Paper-Conference.pdf.

Shengze Cai, Zhicheng Wang, Chryssostomos Chryssostomidis, and George Em Karniadakis. Heat Transfer Prediction With Unknown Thermal Boundary Conditions Using Physics-Informed Neural Networks. In *Volume 3: Computational Fluid Dynamics; Micro and Nano Fluid Dynamics*, pp. V003T05A054, 2020. doi: 10.1115/fedsm2020-20159.

Shengze Cai, Zhicheng Wang, Sifan Wang, Paris Perdikaris, and George Em Karniadakis. Physics-Informed Neural Networks for Heat Transfer Problems. *Journal of Heat Transfer*, 143(6):060801, 2021. doi: 10.1115/1.4050542.

Tianping Chen and Hong Chen. Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems. *IEEE Transactions on Neural Networks*, 6(4):911–917, 1995. doi: 10.1109/72.392253.

Shan Ding, Yongfu Tian, Lang Qin, Hongxiang Ma, and Rui Yang. Physics-informed hierarchical neural operator for solving inverse problem of unsteady heat conduction. *International Journal of Heat and Mass Transfer*, 258:128335, 2026. doi: 10.1016/j.ijheatmasstransfer.2026.128335.

Somdatta Goswami, Katiana Kontolati, Michael D. Shields, and George Em Karniadakis. Deep transfer operator learning for partial differential equations under conditional shift. *Nature Machine Intelligence*, 4(12):1155–1164, 2022. doi: 10.1038/s42256-022-00569-2.

Zhongkai Hao, Chang Su, Songming Liu, Julius Berner, Chengyang Ying, Hang Su, Anima Anandkumar, Jian Song, and Jun Zhu. DPOT: Auto-Regressive Denoising Operator Transformer for Large-Scale PDE Pre-Training. In *Proceedings of the 41st International Conference on Machine*

Learning, volume 235 of *Proceedings of Machine Learning Research*, pp. 17616–17635. PMLR, 2024. URL https://proceedings.mlr.press/v235/hao24d.html.

Zhili He, Futao Ni, Weiguo Wang, and Jian Zhang. A physics-informed deep learning method for solving direct and inverse heat conduction problems of materials. *Materials Today Communications*, 28:102719, 2021. doi: 10.1016/j.mtcomm.2021.102719.

Maximilian Herde, Bogdan Raonić, Tobias Rohner, Roger Käppeli, Roberto Molinaro, Emmanuel de Bézenac, and Siddhartha Mishra. Poseidon: Efficient Foundation Models for PDEs. In *Advances in Neural Information Processing Systems*, volume 37, pp. 72525–72624. Curran Associates, Inc., 2024. doi: 10.52202/079017-2311.

Yuanming Hu, Luke Anderson, Tzu-Mao Li, Qi Sun, Nathan Carr, Jonathan Ragan-Kelley, and Frédo Durand. DiffTaichi: Differentiable Programming for Physical Simulation. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/pdf/6d9976c7113eb4ad907e38be6d4797388ff35a3b.pdf.

Xinchao Jiang, Xin Wang, Ziming Wen, Enying Li, and Hu Wang. Practical uncertainty quantification for space-dependent inverse heat conduction problem via ensemble physics-informed neural networks. *International Communications in Heat and Mass Transfer*, 147:106940, 2023. doi: 10.1016/j.icheatmasstransfer.2023.106940.

Junhyoung Jo, Yeonhwi Jeong, Jinsu Kim, and Jihyung Yoo. Thermal conductivity estimation using Physics-Informed Neural Networks with limited data. *Engineering Applications of Artificial Intelligence*, 137:109079, 2024. doi: 10.1016/j.engappai.2024.109079.

Petr Karnakov, Sergey Litvinov, and Petros Koumoutsakos. Flow reconstruction by multiresolution optimization of a discrete loss with automatic differentiation. *The European Physical Journal E*, 46(7):59, 2023. doi: 10.1140/epje/s10189-023-00313-7.

Petr Karnakov, Sergey Litvinov, and Petros Koumoutsakos. Solving inverse problems in physics by optimizing a discrete loss: Fast and accurate learning without neural networks. *PNAS Nexus*, 3(1): pgae005, 2024. doi: 10.1093/pnasnexus/pgae005.

George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021. doi: 10.1038/s42254-021-00314-5.

Katiana Kontolati, Somdatta Goswami, George Em Karniadakis, and Michael D. Shields. Learning nonlinear operators in latent spaces for real-time predictions of complex dynamics in physical systems. *Nature Communications*, 15(1):5101, 2024. doi: 10.1038/s41467-024-49411-w.

Seid Koric and Diab W. Abueidda. Data-driven and physics-informed deep learning operators for solution of heat conduction equation with parametric heat source. *International Journal of Heat and Mass Transfer*, 203:123809, 2023. doi: 10.1016/j.ijheatmasstransfer.2022.123809.

Nikola Kovachki, Samuel Lanthaler, and Siddhartha Mishra. On Universal Approximation and Error Bounds for Fourier Neural Operators. *Journal of Machine Learning Research*, 22(290):1–76, 2021. URL http://jmlr.org/papers/v22/21-0806.html.

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. *Journal of Machine Learning Research*, 24(89):1–97, 2023. URL http://jmlr.org/papers/v24/21-1524.html.

Samuel Lanthaler, Siddhartha Mishra, and George E Karniadakis. Error estimates for DeepONets: a deep learning framework in infinite dimensions. *Transactions of Mathematics and Its Applications*, 6(1):tnac001, 2022. doi: 10.1093/imatrm/tnac001.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier Neural Operator for Parametric Partial Differential Equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=c8P9NQVtmnO.

Zongyi Li, Daniel Zhengyu Huang, Burigede Liu, and Anima Anandkumar. Fourier Neural Operator with Learned Deformations for PDEs on General Geometries. *Journal of Machine Learning Research*, 24(388):1–26, 2023. URL http://jmlr.org/papers/v24/23-0064.html.

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. doi: 10.1145/3648506.

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, 2021. doi: 10.1038/s42256-021-00302-5.

Lu Lu, Xuhui Meng, Shengze Cai, Zhiping Mao, Somdatta Goswami, Zhongqiang Zhang, and George Em Karniadakis. A comprehensive and fair comparison of two neural operators (with practical extensions) based on FAIR data. *Computer Methods in Applied Mechanics and Engineering*, 393:114778, 2022. doi: 10.1016/j.cma.2022.114778.

Siddhartha Mishra and Roberto Molinaro. Estimates on the generalization error of physics-informed neural networks for approximating a class of inverse problems for PDEs. *IMA Journal of Numerical Analysis*, 42(2):981–1022, 2022. doi: 10.1093/imanum/drab032.

Roberto Molinaro, Yunan Yang, Björn Engquist, and Siddhartha Mishra. Neural inverse operators for solving pde inverse problems. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 25105–25139. PMLR, 2023.

Hyeonbin Moon, Songho Lee, Wabi Demeke, Byungki Ryu, and Seunghwa Ryu. Physics-informed neural operators for generalizable and label-free inference of temperature-dependent thermoelectric properties. *npj Computational Materials*, 11(1):272, 2025. doi: 10.1038/s41524-025-01769-1.

Vivek Oommen and Balaji Srinivasan. Solving Inverse Heat Transfer Problems Without Surrogate Models: A Fast, Data-Sparse, Physics Informed Neural Network Approach. *Journal of Computing and Information Science in Engineering*, 22(4):041012, 2022. doi: 10.1115/1.4053800.

W. J. Parker, R. J. Jenkins, C. P. Butler, and G. L. Abbott. Flash Method of Determining Thermal Diffusivity, Heat Capacity, and Thermal Conductivity. *Journal of Applied Physics*, 32(9):1679–1684, 1961. doi: 10.1063/1.1728417.

Wenkai Qiu, Haolong Chen, and Huanlin Zhou. A novel hybrid deep learning algorithm for estimating temperature-dependent thermal conductivity in transient heat conduction problems. *International Communications in Heat and Mass Transfer*, 164:108871, 2025. doi: 10.1016/j.icheatmasstransfer.2025.108871.

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

Yuanbin Shang, Huaiguo Ban, and Donghuan Liu. Simultaneous identification of boundary heat flux and thermal conductivity in inverse heat conduction problems using physics-informed neural networks. *Thermal Science and Engineering Progress*, 65:103905, 2025. doi: 10.1016/j.tsep.2025.103905.

Hong Tang, Alexander Melnikov, MingRui Liu, Stefano Sfarra, Hai Zhang, and Andreas Mandelis. Physics informed neural networks for solving inverse thermal wave coupled boundary-value problems. *International Journal of Heat and Mass Transfer*, 245:126985, 2025. doi: 10.1016/j.ijheatmasstransfer.2025.126985.

Alasdair Tran, Alexander Mathews, Lexing Xie, and Cheng Soon Ong. Factorized Fourier Neural Operators. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=tmIiMPl4IPa.

Kiwon Um, Robert Brand, Yun (Raymond) Fei, Philipp Holl, and Nils Thuerey. Solver-in-the-Loop: Learning from Differentiable Physics to Interact with Iterative PDE-Solvers. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6111–6122. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/43e4e6a6f341e00671e123714de019a8-Paper.pdf.

Sifan Wang, Hanwen Wang, and Paris Perdikaris. Learning the solution operator of parametric partial differential equations with physics-informed DeepONets. *Science Advances*, 7(40):eabi8605, 2021. doi: 10.1126/sciadv.abi8605.

Sean Wu, Shamik Basu, Tim Brödermann, Luc Van Gool, and Christos Sakaridis. Pbr-nerf: Inverse rendering with physics-based neural fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10974–10984, 2025.

Liu Yang, Xuhui Meng, and George Em Karniadakis. B-PINNs: Bayesian physics-informed neural networks for forward and inverse PDE problems with noisy data. *Journal of Computational Physics*, 425:109913, 2021. doi: 10.1016/j.jcp.2020.109913.

Jeremy Yu, Lu Lu, Xuhui Meng, and George Em Karniadakis. Gradient-enhanced physics-informed neural networks for forward and inverse PDE problems. *Computer Methods in Applied Mechanics and Engineering*, 393:114823, 2022. doi: 10.1016/j.cma.2022.114823.

Martin Zálešák, Lubomír Klimeš, Pavel Charvát, Matouš Cabalka, Jakub Kůdela, and Tomáš Mauder. Solution approaches to inverse heat transfer problems with and without phase changes: A state-of-the-art review. *Energy*, 278:127974, 2023. doi: 10.1016/j.energy.2023.127974.

Bo Zhang and Zhen Li. Data-driven and physics-constrained learning for inferring physical parameters and solving forward problems of heat conduction. *Applied Thermal Engineering*, 264:125534, 2025. doi: 10.1016/j.applthermaleng.2025.125534.

Tao Zhong, Yixun Hu, Dongzhe Zheng, Aditya Sood, and Christine Allen-Blanchette. Neural Field Thermal Tomography: A Differentiable Physics Framework for Non-Destructive Evaluation, 2026. arXiv preprint arXiv:2603.11045.

Hang Zhou, Haixu Wu, Haonan Shangguan, Yuezhou Ma, Huikun Weng, Jianmin Wang, and Mingsheng Long. Transolver-3: Scaling Up Transformer Solvers to Industrial-Scale Geometries, 2026. arXiv preprint arXiv:2602.04940.

Navid Zobeiry and Keith D. Humfeld. A physics-informed machine learning approach for solving heat transfer equation in advanced manufacturing and engineering applications. *Engineering Applications of Artificial Intelligence*, 101:104232, 2021. doi: 10.1016/j.engappai.2021.104232.

A MatTherm: A Benchmark for Material Identification and Thermal Evolution Prediction

MatTherm evaluates material identification from temperature observations and thermal evolution prediction using the inferred conductivity. Each material is defined by a temperature-dependent conductivity function $k(T)$ and simulated under different geometries, initial temperatures, and heat sources. The 2D and 3D datasets share the same conductivity library, and models are trained separately for each dimension. Table 4 summarizes their scale and composition.

A.1 Dataset Construction

Conductivity functions and material splits. The synthetic conductivity library covers five response archetypes: monotonic increase, monotonic decrease, a single peak, a single valley, and multiple turning points. These archetypes describe the shape of $k(T)$ over the temperature range. The library comprises 15 functional subfamilies in total, capturing variation in temperature dependence within these response patterns (Figure 7). We sample function parameters and a multiplicative scale within each subfamily, retaining curves with finite, positive conductivity values. Each function is stored at 256 uniformly spaced temperatures over $[200, 1000]$ K. The training library covers all five response archetypes and is approximately balanced across the 15 functional subfamilies. The training, validation, and test partitions contain distinct materials. For the main-text experiments, seen materials belong to the thermal training set, and unseen materials come from the conductivity test partition.

Geometries and heat sources. Each dimension contains 13 geometry families and 15 source families, with the training and held-out splits listed in Table 4. Geometries include regular, curved, non-convex, and perforated domains, with variation in size, position, orientation, and shape (Figure 8). Training sources include localized and broad Gaussian heating, boundary and interior strips or slabs, multiple hotspots, and temporal pulses. Held-out sources are ring, diagonal, and moving-Gaussian patterns in 2D, and spherical-shell, diagonal-tube, and moving-Gaussian patterns in 3D. All sources are nonnegative and confined to the material domain.

Table 4: **MatTherm overview.** Dataset size and family splits in 2D and 3D. Both dimensions use the same conductivity library.

Component	2D	3D
Spatial resolution	64×64	$32 \times 32 \times 32$
Training materials / trajectories	1,000 / 12,000	1,000 / 12,000
Standard / long sequence length	40 / 80 frames	40 / 80 frames
Training / held-out geometry families	8 / 5	8 / 5
Training / held-out source families	12 / 3	12 / 3
Conductivity functions: train / validation / test	5,000 / 500 / 500	
Response archetypes / functional subfamilies	5 / 15	
Temperature range / points per conductivity curve	$[200, 1000]$ K / 256	
Thermal training temperature range	$[200, 600]$ K	

Thermal trajectories. We simulate equation 1 using a conservative finite-volume scheme with harmonic interface conductivity and explicit Euler time integration. The simulation uses $a = 10^{-4}$, $\beta = 0.1$, $T_{\text{amb}} = 300$ K, and $\Delta t = 0.002$ s. Conductivity is interpolated at the current temperature during each time step. Zero normal heat flux is imposed on the outer boundary and on internal holes or cavities. Cells outside the material remain at ambient temperature.

Initial temperatures combine a sampled base level with weak noise and a smooth spatial perturbation. Source amplitudes are adjusted to keep training temperatures within $[200, 600]$ K. We retain trajectories with a temperature span of at least 35 K and spatial temperature variation inside the material. Standard and long trajectories contain 40 and 80 evolution frames, respectively; the initial field is stored separately. Observation and prediction windows are specified in Appendix C.

Multiple conditions for each training material. Each dimension uses 1,000 training materials, with 12 trajectories per material. Every material appears under all 12 training source families, with

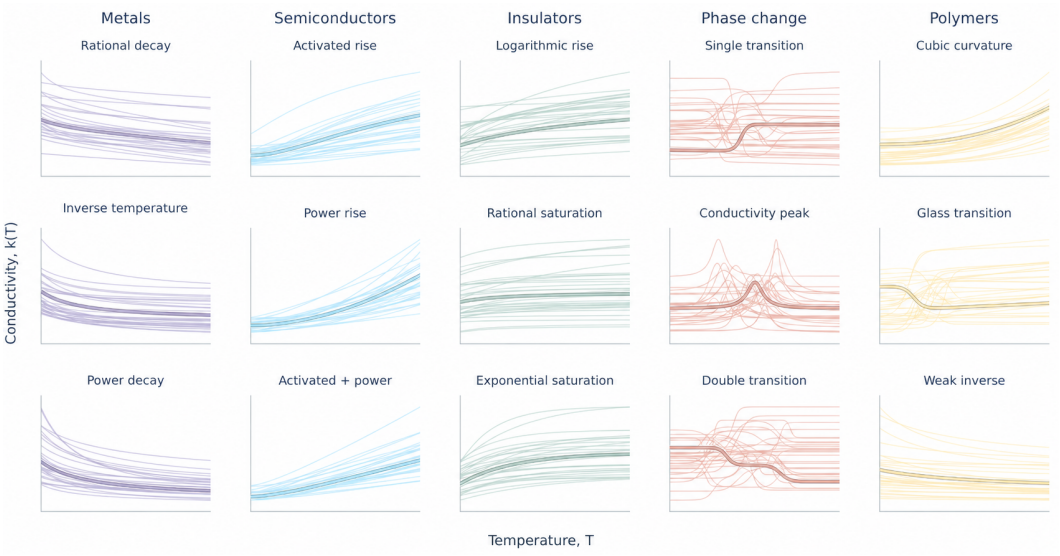

Figure 7: **Constitutive-law diversity.** Sampled conductivity laws illustrate diverse temperature-response patterns and variation within functional subfamilies.

geometry assignments covering the eight training geometry families. These groups support learning a consistent material representation across conditions through the objective in equation 7. At test time, each material identification uses one observed sequence.

A.2 Sample Contents

Temperature and physical information. Each sample stores the clean temperature trajectory, noisy temperature observations, initial field, reference heat-source field, and a link to the material's conductivity function. Geometry is represented by a domain mask, the fraction of each cell occupied by material, and a signed distance field. The sample also records geometry and source parameters, an observation mask, and the grid spacing, time step, and known physical coefficients. Material identifiers connect trajectories generated from the same conductivity function.

Observations. We add Gaussian noise with a standard deviation sampled uniformly from $[0.01, 0.05]$ K. Each scene also has a Gaussian multiplicative gain with mean 1 and standard deviation 0.001, and a Gaussian additive bias with mean 0 and standard deviation 0.15 K. Each temperature entry is independently dropped with probability 5×10^{-4}. The observation mask identifies valid measurements for reconstruction losses, and clean trajectories provide the temperature-prediction targets.

Use in training and identification. Stage I uses training conductivity functions to construct the shared basis and reference material representations. Stage II learns to predict material representations and heat sources from temperature observations. Reference material representations and heat-source fields provide supervision, and temperature reconstruction provides simulation feedback (Section 3.2). The shared material structure and heating patterns learned during training support identification of unseen materials. For a test scene, identification uses one temperature observation sequence, including its initial field, together with the known geometry and boundary conditions. The model predicts both $\widehat{k}(T)$ and the unknown heat source $\widehat{q}(\mathbf{x}, t)$ from these inputs.

A.3 Evaluation Protocols

Prediction under original and new conditions. Table 5 summarizes the evaluations reported in the main text. Under the original operating conditions, the simulator starts from the temperature field at the beginning of the prediction interval and uses the inferred conductivity and predicted source.

Table 5: **Evaluation protocols.** Changes in geometry and heat source are relative to the identification scene. Each transfer prediction uses the conductivity inferred from one observed sequence.

Protocol	Materials	Evaluation target
Long-horizon prediction	500 seen per dimension	Future temperature evolution in 2D and 3D.
Source change	500 unseen in 2D	Temperature evolution under a new source.
Geometry change	500 unseen in 2D	Temperature evolution in a new geometry.
Joint change	500 unseen in 2D	Temperature evolution with both geometry and source changed.
Conductivity recovery	3 seen and 3 unseen	Conductivity accuracy and identification time.



Figure 8: **Geometry diversity.** Training and held-out families cover different boundary shapes and topologies, with variation within each family.

For a new operating condition, the conductivity inferred from one observed sequence is reused with the target geometry, prescribed source, and initial field. The material is unchanged across the identification and target scenes, and the inferred conductivity is held fixed during transfer. Predicted temperatures are compared with the clean target trajectory.

Zero-shot and few-shot evaluation. Zero-shot identification uses one forward pass without per-instance optimization. Few-shot adaptation uses selected frames already included in the identification observations, with the shared basis and simulator fixed. Both modes use the same input observations and prediction targets.

Metrics. Temperature prediction is measured by RMSE in kelvin and response-normalized relative L_2 error. PDE residuals and energy-balance errors measure physical consistency. Conductivity recovery is measured by relative L_2 error over the full evaluation temperature range (Full) and the range covered by the identification observations (Obs.), together with per-instance identification time. For the conductivity comparison in Table 3, PDE residuals are evaluated on transfer trajectories using the reference conductivity. Results are averaged over the materials listed in Table 5.

B Additional Experiments

B.1 Error Evolution and Tail Performance in Long-Horizon Prediction

Extending Table 1 in the main text, we examine long-horizon prediction from two perspectives: accuracy over the prediction horizon and the distribution of errors across materials. We use the same 2D and 3D evaluation settings, with 500 materials per dimension, and compute all temperature errors against clean simulation ground truth. In addition to the overall RMSE, we report RMSE-AUC, RMSE over the final quarter of the prediction horizon (FQ-RMSE), the 95th percentile of per-material RMSE (P95), and the mean RMSE over the worst-performing 5% of materials (CVaR95). RMSE-AUC is computed as the temporal average of per-frame RMSE. FQ-RMSE measures accuracy toward the end of the prediction horizon, whereas P95 and CVaR95 characterize the tail of the error distribution across materials. All metrics are expressed in kelvin, and lower values are better.

Table 6: Detailed errors for 2D long-horizon prediction over 500 materials. All metrics are in kelvin; lower is better. Best results are shown in bold.

Method	RMSE	RMSE-AUC	FQ-RMSE	P95	CVaR95
F-FNO	3.82e−1	3.76e−1	4.27e−1	8.89e−1	1.08e+0
DPOT	8.45e−1	7.81e−1	1.15e+0	2.64e+0	3.35e+0
Poseidon	7.00e−1	6.22e−1	1.00e+0	1.42e+0	4.22e+0
MemNO	4.00e−1	3.89e−1	4.64e−1	9.16e−1	1.30e+0
Transolver3	8.88e+0	7.97e+0	1.31e+1	1.41e+1	1.49e+1
UniMaNO	**1.12e−1**	**1.11e−1**	**1.25e−1**	**3.09e−1**	**5.00e−1**

Table 7: Detailed errors for 3D long-horizon prediction over 500 materials. All metrics are in kelvin; lower is better. Best results are shown in bold.

Method	RMSE	RMSE-AUC	FQ-RMSE	P95	CVaR95
F-FNO	4.06e−1	3.95e−1	4.67e−1	9.25e−1	1.10e+0
DPOT	1.88e+0	1.75e+0	2.41e+0	3.57e+0	9.72e+0
Poseidon	5.82e+0	5.39e+0	8.30e+0	7.70e+0	8.23e+0
MemNO	3.80e−1	3.74e−1	4.12e−1	8.27e−1	9.85e−1
Transolver3	4.90e+0	4.60e+0	6.82e+0	6.99e+0	7.53e+0
UniMaNO	**2.61e−1**	**2.60e−1**	**2.75e−1**	**6.55e−1**	**9.08e−1**

Maintaining accuracy at late prediction times. UniMaNO achieves the lowest RMSE-AUC and FQ-RMSE in both 2D and 3D (Tables 6 and 7). Relative to the strongest neural operator baseline in each dimension, it reduces RMSE-AUC and FQ-RMSE by 70.5% and 70.8% in 2D, and by 30.5% and 33.3% in 3D, respectively. Thus, the overall RMSE advantage reported in the main text persists into the final quarter of the prediction horizon. In 3D, the reduction in FQ-RMSE slightly exceeds the reduction in overall RMSE (31.1%), supporting UniMaNO's ability to maintain prediction accuracy at later times. These results are consistent with its design of driving physical evolution using identified material functions.

Tail errors across materials. UniMaNO's advantage in average accuracy also extends to the tail of the per-material error distribution. In 2D, it reduces P95 and CVaR95 by 65.2% and 53.8%, respectively, relative to F-FNO. In 3D, the corresponding reductions relative to MemNO are 20.8% and 7.8%. Lower average errors therefore coincide with lower high-percentile errors and lower mean errors among the worst-performing cases. However, the smaller improvement in 3D CVaR95, compared with overall RMSE and P95, indicates that the advantage narrows for the most challenging materials. This identifies a remaining limitation: although UniMaNO performs best in average accuracy, late-horizon accuracy, and tail metrics, the most difficult 3D cases remain an important target for further improvement.

Table 8: Thermal evolution prediction for seen materials under heat-source changes. Results are averaged over 500 materials; all methods use zero-shot inference.

Method	RMSE	L_2	AUC	FQ-RMSE	PDE	Energy
F-FNO	9.66e−01	4.03e−02	8.29e−01	1.46e+00	2.85e−01	7.13e−02
DPOT	3.45e+00	1.37e−01	2.95e+00	5.19e+00	3.97e−01	4.35e−01
Poseidon	3.90e+00	1.52e−01	3.27e+00	6.04e+00	3.95e−01	3.79e−01
MemNO	9.54e−01	3.76e−02	8.16e−01	1.44e+00	2.82e−01	7.49e−02
Transolver3	2.04e+01	8.15e−01	1.86e+01	2.64e+01	1.42e+00	9.25e−01
UniMaNO	5.91e−02	2.45e−03	4.88e−02	9.36e−02	3.66e−03	7.81e−06

Table 9: Thermal evolution prediction for seen materials under geometry changes. Results are averaged over 500 materials; all methods use zero-shot inference. † marks extreme prediction errors retained in the reported statistics.

Method	RMSE	L_2	AUC	FQ-RMSE	PDE	Energy
F-FNO	7.77e−01	5.99e−02	7.05e−01	1.07e+00	4.85e−01	5.97e−02
DPOT†	7.71e+11	4.54e+10	1.68e+11	1.49e+12	7.42e−01	6.87e−01
Poseidon	2.88e+00	1.78e−01	2.43e+00	4.42e+00	6.30e−01	3.35e−01
MemNO	1.57e+00	9.95e−02	1.27e+00	2.48e+00	5.62e−01	2.25e−01
Transolver3	1.36e+01	1.05e+00	1.25e+01	1.75e+01	1.41e+00	1.25e+00
UniMaNO	5.49e−02	3.62e−03	4.71e−02	8.35e−02	8.47e−03	1.71e−04

B.2 Additional Cross-Condition Generalization Results

Extending Section 4.3, we distinguish condition shifts during material identification from those during subsequent prediction. The former evaluates identification from temperature observations under new heat-source or geometry conditions; the latter evaluates reuse of an inferred material law for prediction under new conditions. Whether a material was seen during training is treated as a separate evaluation dimension. All results below use zero-shot inference without test-time parameter updates, and our method is denoted as UniMaNO. Lower values are better for all metrics. In the temperature prediction tables, L_2 denotes response-normalized relative L_2 error, AUC denotes time-averaged RMSE over the prediction window, and FQ-RMSE denotes RMSE over its final quarter. RMSE, AUC, and FQ-RMSE are reported in K.

Generalization under prediction-condition shifts. We extend Table 2 with seen-material comparisons and additional metrics for temporal prediction error and physical consistency. Tables 8–13 report heat-source, geometry, and joint changes for seen and unseen materials, with 500 materials per setting. The RMSE and L_2 values for unseen materials match the corresponding zero-shot results in Table 2.

Across the three condition changes, the RMSE difference between seen and unseen materials remains below $4.00\text{e}{-03}$ K, with comparable PDE residuals and energy balance errors. This indicates limited sensitivity to whether a material appeared during training in these evaluations. For unseen materials, UniMaNO achieves RMSEs of $5.63\text{e}{-02}$, $5.13\text{e}{-02}$, and $6.82\text{e}{-02}$ K under heat-source, geometry, and joint changes, respectively, reducing error by 94.1%, 93.4%, and 94.0% relative to the best neural operator baseline in each setting. UniMaNO also achieves lower AUC, FQ-RMSE, PDE residuals, and energy balance errors. Under joint changes, its FQ-RMSE remains at $1.12\text{e}{-01}$ K for seen materials and $1.07\text{e}{-01}$ K for unseen materials. These results support reuse of the shared material representation under new prediction conditions and show a consistent prediction advantage across the evaluated material and condition combinations.

Robustness to identification-condition shifts. We further examine condition changes in the observations used for material identification. This experiment uses 500 seen materials and reports material-law recovery and downstream prediction errors under a control condition, a heat-source shift, and a geometry shift (Tables 14 and 15). These changes affect the identification inputs and are evaluated separately from the prediction-condition shifts in Table 2.

Table 10: Thermal evolution prediction for seen materials under joint heat-source and geometry changes. Results are averaged over 500 materials; all methods use zero-shot inference.

Method	RMSE	L_2	AUC	FQ-RMSE	PDE	Energy
F-FNO	1.15e+00	5.80e−02	9.86e−01	1.73e+00	3.46e−01	7.12e−02
DPOT	5.81e+00	2.88e−01	4.78e+00	9.18e+00	6.29e−01	5.29e−01
Poseidon	5.89e+00	3.15e−01	4.38e+00	9.53e+00	5.38e−01	4.15e−01
MemNO	1.62e+00	8.15e−02	1.31e+00	2.61e+00	3.84e−01	1.54e−01
Transolver3	1.83e+01	8.73e−01	1.67e+01	2.37e+01	1.42e+00	9.89e−01
UniMaNO	7.14e−02	3.88e−03	5.92e−02	1.12e−01	6.17e−03	1.05e−04

Table 11: Thermal evolution prediction for unseen materials under heat-source changes. Results are averaged over 500 materials; all methods use zero-shot inference.

Method	RMSE	L_2	AUC	FQ-RMSE	PDE	Energy
F-FNO	9.50e−01	3.96e−02	8.16e−01	1.44e+00	2.54e−01	6.94e−02
DPOT	3.42e+00	1.35e−01	2.92e+00	5.13e+00	3.70e−01	4.35e−01
Poseidon	3.95e+00	1.55e−01	3.31e+00	6.10e+00	3.71e−01	3.79e−01
MemNO	9.49e−01	3.74e−02	8.11e−01	1.44e+00	2.52e−01	7.52e−02
Transolver3	2.04e+01	8.12e−01	1.86e+01	2.64e+01	1.42e+00	9.25e−01
UniMaNO	5.63e−02	2.33e−03	4.62e−02	8.95e−02	3.46e−03	7.03e−06

Under an identification geometry shift, downstream RMSE is $9.94\text{e}{-02}$ K, an increase of $4.03\text{e}{-02}$ K over the control. An identification heat-source shift yields an RMSE of $4.10\text{e}{-01}$ K. Although the heat-source shift causes a larger increase, downstream RMSE under the geometry and heat-source shifts remains 89.6% and 57.0% lower, respectively, than the best neural operator reference result. The corresponding PDE residuals are $6.05\text{e}{-03}$ and $2.49\text{e}{-02}$. UniMaNO thus retains a downstream prediction accuracy advantage when identification conditions change. Meanwhile, the increased material-law error under the heat-source shift shows that accurate downstream prediction does not necessarily imply accurate law recovery across all evaluated temperatures. Neural operator results serve as a common downstream reference and are identical across the three identification conditions in the available records; this comparison therefore concerns prediction accuracy, not relative degradation under identical identification perturbations.

B.3 Learned Priors and Cross-Condition Validity

Context from inverse problems. Recovering physical properties from limited observations can involve ambiguities when multiple physical factors jointly determine the measured response. In inverse rendering, PBR-NeRF introduces physics-based priors to constrain joint material and illumination estimation (Wu et al., 2025). In thermoelectric property inference, physical equations are combined with prescribed material information to constrain the recovery of unknown properties (Moon et al., 2025). Neural inverse operators further provide a framework for learning inverse mappings across instances (Molinaro et al., 2023). These studies illustrate the importance of specifying the available information and incorporating appropriate constraints when learning inverse estimators.

Learned constraints in UniMaNO. UniMaNO combines complementary constraints for joint conductivity and heat-source estimation. The shared log-POD representation restricts conductivity prediction to a low-dimensional family constructed from training conductivity curves. Material supervision anchors predicted coefficients to reference representations, while source supervision trains the model to account explicitly for external heating. Same-material consistency penalizes representation differences across operating conditions, encouraging material coefficients to capture constitutive behavior rather than scene-specific heating. Differentiable simulation additionally penalizes disagreement between observed temperatures and the evolution produced by the predicted conductivity and heat source. Identification therefore uses both cross-material structure and scene-level physical feedback, rather than temperature reconstruction alone. In zero-shot inference, the

Table 12: Thermal evolution prediction for unseen materials under geometry changes. Results are averaged over 500 materials; all methods use zero-shot inference. † marks extreme prediction errors retained in the reported statistics.

Method	RMSE	L_2	AUC	FQ-RMSE	PDE	Energy
F-FNO	7.74e−01	6.00e−02	7.03e−01	1.07e+00	4.58e−01	5.94e−02
DPOT†	4.41e+13	2.59e+12	9.55e+12	8.55e+13	7.22e−01	6.86e−01
Poseidon	2.89e+00	1.79e−01	2.43e+00	4.44e+00	6.01e−01	3.34e−01
MemNO	1.56e+00	9.95e−02	1.27e+00	2.47e+00	5.37e−01	2.22e−01
Transolver3	1.36e+01	1.04e+00	1.25e+01	1.75e+01	1.41e+00	1.25e+00
UniMaNO	5.13e−02	3.44e−03	4.41e−02	7.80e−02	8.15e−03	1.64e−04

Table 13: Thermal evolution prediction for unseen materials under joint heat-source and geometry changes. Results are averaged over 500 materials; all methods use zero-shot inference.

Method	RMSE	L_2	AUC	FQ-RMSE	PDE	Energy
F-FNO	1.13e+00	5.71e−02	9.72e−01	1.71e+00	3.29e−01	7.12e−02
DPOT	5.83e+00	2.91e−01	4.79e+00	9.23e+00	6.20e−01	5.29e−01
Poseidon	5.05e+00	2.43e−01	4.18e+00	7.89e+00	5.27e−01	4.15e−01
MemNO	1.59e+00	7.99e−02	1.29e+00	2.57e+00	3.66e−01	1.52e−01
Transolver3	1.83e+01	8.73e−01	1.66e+01	2.37e+01	1.42e+00	9.89e−01
UniMaNO	6.82e−02	3.57e−03	5.64e−02	1.07e−01	5.65e−03	1.05e−04

trained model estimates conductivity and heat source from one temperature sequence and the known physical setting, without heat-source measurements or per-instance optimization.

Validation beyond reconstruction. We assess whether the inferred conductivity remains useful beyond the sequence used for identification. In cross-condition prediction, the conductivity is held fixed and combined with the target geometry, initial temperature, and prescribed heat source, rather than with the source estimated for the identification scene. This tests its predictive value outside the original conductivity–source combination. In the three 2D cross-condition settings, mean zero-shot RMSE is 0.0563, 0.0513, and 0.0682 K under source, geometry, and joint changes, respectively, with 500 unseen materials per setting. These prediction-condition shifts are distinct from changes in the observations used for identification, which we evaluate separately. The identification-shift experiments show that full-range conductivity recovery and downstream prediction can respond differently to changes in heating conditions. Together, these evaluations provide empirical evidence for the reusability of the inferred material functions within the tested materials and operating conditions.

B.4 Reconstruction Accuracy of the Shared Material Representation

A key requirement of the shared material representation is that its low-dimensional parameterization preserve material-dependent conductivity variations without introducing substantial basis truncation error. To assess this requirement, we fix the rank-32 Log-POD basis constructed in Stage I, project the ground-truth conductivity curves of test materials onto this basis in log space, and reconstruct the conductivity functions through the decoder. Evaluation covers the full temperature range of 200–1000 K. Since the material coefficients are obtained directly from the ground-truth curves, this experiment isolates representation error from errors in estimating coefficients from observations.

The rank-32 shared basis achieves a mean reconstruction error of 9.74×10^{-5}, with an upper 95% confidence bound of 1.27×10^{-4}. Beyond this low average error, the maximum relative error remains below 0.44%, indicating that the aggregate accuracy does not conceal severe representation failures on individual test materials. Thus, within the evaluated test set, a single fixed basis provides accurate reconstruction for both typical and worst-case materials.

These results support parameterizing material identification as coefficient prediction in a shared function space. Rather than learning an independent function representation for each material, the model can preserve conductivity variations through 32 material-specific coefficients. The small

Table 14: Material-law recovery by UniMaNO under identification-condition shifts. Each condition includes 500 seen materials. Intervals are 95% confidence intervals for the mean relative L_2 error.

Identification condition	Law rel. L_2	95% CI	Log-RMSE
Control	3.63e−01	[3.27e−01, 4.05e−01]	3.61e−01
Source shift	2.74e+00	[2.09e+00, 3.50e+00]	7.99e−01
Geometry shift	4.22e−01	[3.90e−01, 4.57e−01]	4.21e−01

Table 15: Downstream thermal evolution prediction under identification-condition shifts. Each condition includes 500 seen materials. Neural operator results are identical across the three conditions in this experiment and are listed once as a common reference. UniMaNO results are reported separately for each identification condition.

Method	RMSE	L_2	AUC	FQ-RMSE	PDE	Energy
F-FNO	9.66e−01	4.03e−02	8.29e−01	1.46e+00	2.85e−01	7.13e−02
DPOT	3.45e+00	1.37e−01	2.95e+00	5.19e+00	3.97e−01	4.35e−01
Poseidon	3.90e+00	1.52e−01	3.27e+00	6.04e+00	3.95e−01	3.79e−01
MemNO	9.54e−01	3.76e−02	8.16e−01	1.44e+00	2.82e−01	7.49e−02
Transolver3	2.04e+01	8.15e−01	1.86e+01	2.64e+01	1.42e+00	9.25e−01
UniMaNO (control)	5.91e−02	2.45e−03	4.88e−02	9.36e−02	3.66e−03	7.81e−06
UniMaNO (source shift)	4.10e−01	1.57e−02	3.32e−01	6.62e−01	2.49e−02	7.48e−06
UniMaNO (geometry shift)	9.94e−02	3.97e−03	8.14e−02	1.58e−01	6.05e−03	7.36e−06

Table 16: Projection reconstruction errors of the rank-32 shared Log-POD basis on test materials. Errors are measured using relative L_2 over the full temperature range of 200–1000 K; lower values are better.

Statistic	Relative L_2 error ↓
Mean	9.74×10^{-5}
Median	2.25×10^{-5}
Maximum	4.37×10^{-3}
95% confidence interval for the mean	$[7.21 \times 10^{-5}, 1.27 \times 10^{-4}]$

projection error also provides a reference for separating two questions: whether the shared basis can represent a material law, and whether its coefficients can be accurately recovered from temperature observations. Table 16 evaluates the former, while the material identification and thermal prediction experiments in the main text assess the latter and its physical relevance.

B.5 Additional Analysis of Loss Ablations

Table 17: **Zero-shot loss ablations.** All metrics ↓; bold denotes the best result in each column.

	Cross-condition (seen materials, heat-source change)			**Long-horizon**	
Variant	RMSE (K)	L2	PDE	Tail RMSE (K)	Energy
Without $\mathcal{L}_{\text{sim}}$	5.99e−02	2.78e−03	3.74e−03	1.51e−01	3.03e−05
Without $\mathcal{L}_{\text{inv}}$	5.95e−02	2.46e−03	3.72e−03	1.37e−01	3.00e−05
Full model	**5.91e−02**	**2.45e−03**	**3.66e−03**	**1.25e−01**	**2.99e−05**

We examine the contributions of $\mathcal{L}_{\text{sim}}$ and $\mathcal{L}_{\text{inv}}$ by removing each loss separately and evaluating the resulting models under cross-condition and long-horizon settings (Table 17). Cross-condition performance is assessed using RMSE, L2 error, and the PDE residual. For long-horizon prediction, Tail RMSE measures temperature error over the final quarter of the prediction window, while Energy

is evaluated over the full window. All temperature errors are computed against clean simulation ground truth.

The full model achieves the lowest values across all reported metrics. Removing either loss produces a modest increase in cross-condition RMSE, but a more pronounced degradation in late-horizon accuracy. Specifically, removing $\mathcal{L}_{\text{sim}}$ increases Tail RMSE from 0.125 to $0.151\,\text{K}$, a relative increase of 20.8%, while removing $\mathcal{L}_{\text{inv}}$ increases it to $0.137\,\text{K}$, a relative increase of 9.6%. Both ablations also slightly increase the PDE residual and energy error. These results support retaining both losses, with $\mathcal{L}_{\text{sim}}$ making the larger contribution to late-horizon accuracy in this evaluation.

B.6 Additional Qualitative Results

We provide additional qualitative comparisons on the 2D and 3D datasets. Figure 9 compares predictions across different geometries and temperature patterns, while Figures 10 and 11 illustrate their temporal evolution. Figures 12 and 13 further show whole and cutaway views of two 3D cases, revealing both exterior and interior temperature distributions. Across these examples, our method closely matches the ground truth and maintains low prediction errors across space and time.

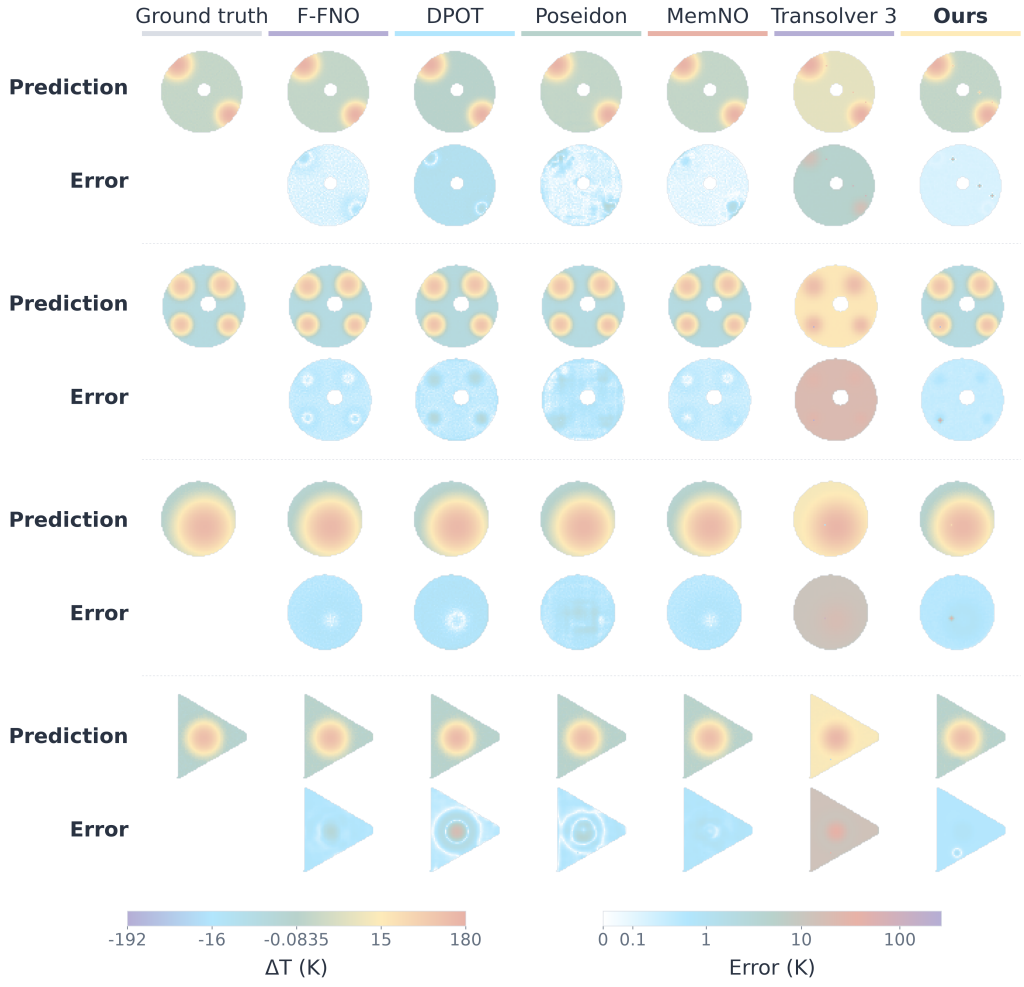

Figure 9: **Qualitative comparison on the 2D dataset.** Four examples illustrate different geometries and spatial temperature patterns. For each example, the upper row shows ground-truth and predicted temperature variations ΔT, and the lower row shows the corresponding absolute errors. Our method closely reproduces the spatial temperature distributions with reduced local errors.

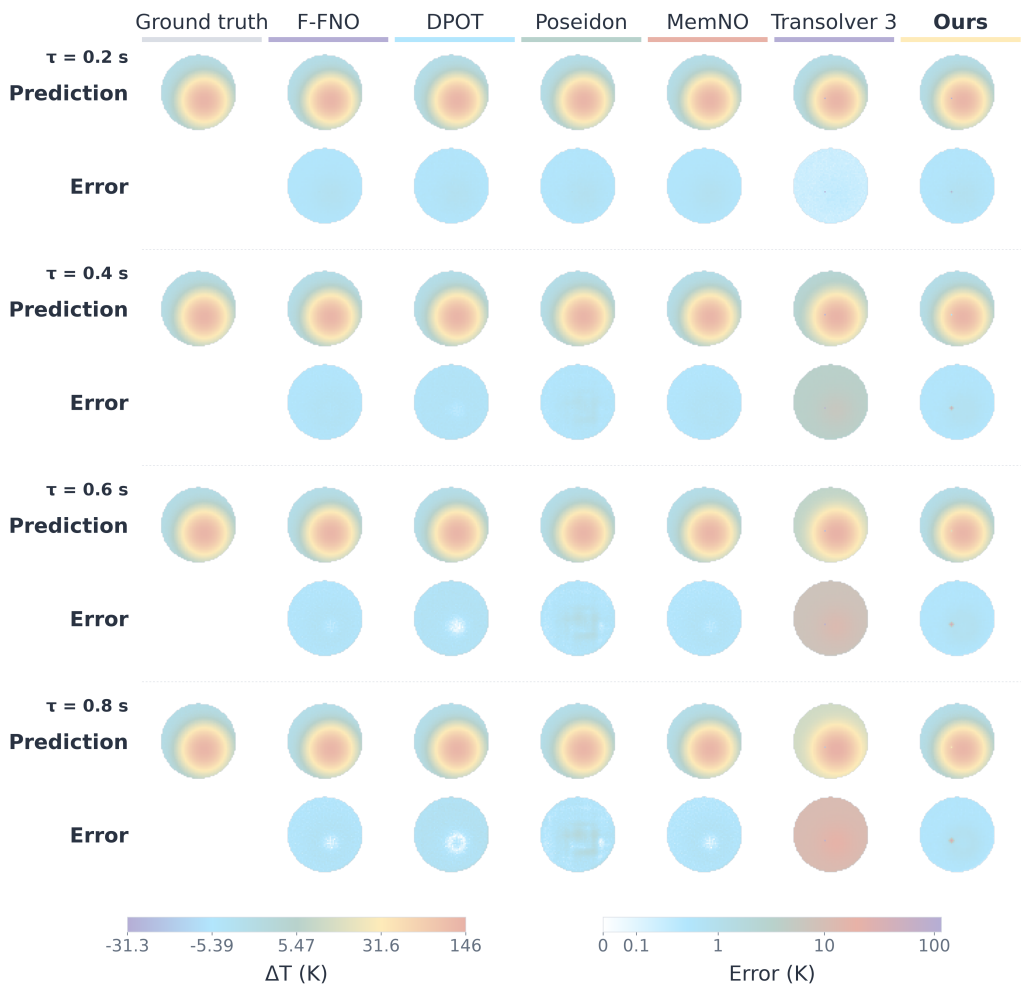

Figure 10: **Temporal evolution of 2D temperature predictions.** Each pair of rows shows the temperature change relative to the ground-truth field at the observation time, $T(t) - T_{\mathrm{GT}}(t_{\mathrm{obs}})$, and the corresponding absolute prediction error. Results at four successive times show that our method captures the evolving localized cooling pattern while maintaining low errors.

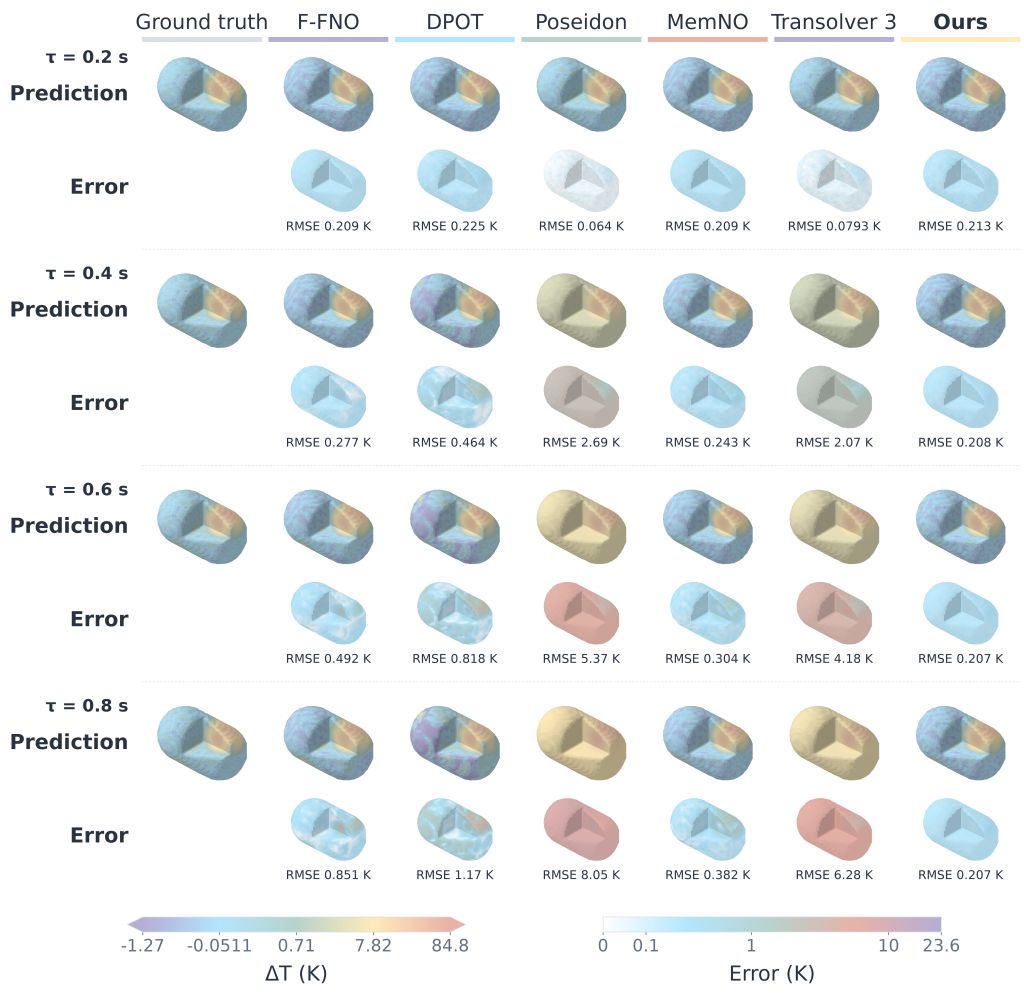

Figure 11: **Temporal evolution of 3D temperature predictions.** Cutaway views show the temperature change relative to the ground-truth field at the observation time, $T(t) - T_{\mathrm{GT}}(t_{\mathrm{obs}})$, together with the corresponding absolute prediction errors at four successive times. Our method closely tracks the cooling dynamics across both the exterior and the exposed interior.

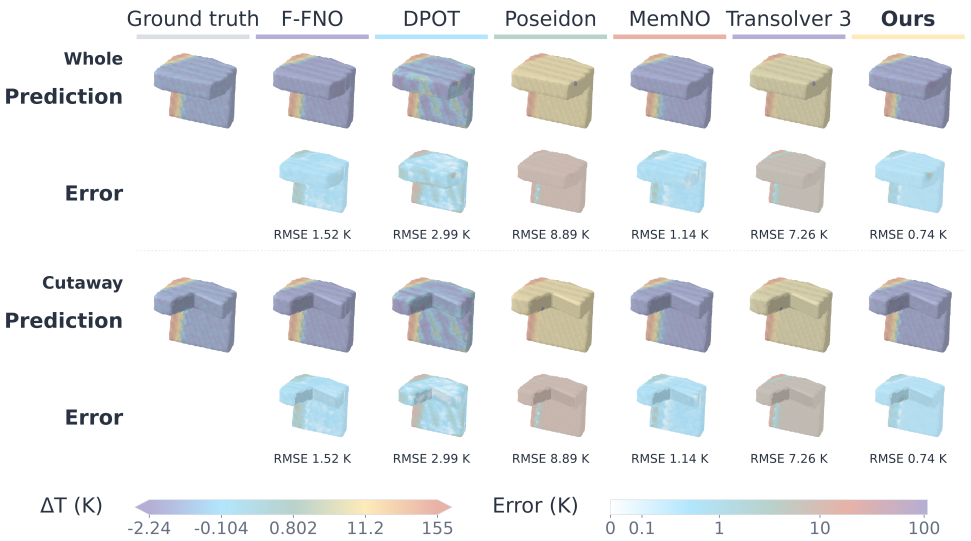

Figure 12: **Qualitative comparison for 3D Case 1.** Whole and cutaway views show ground-truth and predicted temperature variations ΔT, with absolute errors displayed below each prediction. The cutaway reveals temperature variations within the spherical domain. Our method closely reproduces the spatial temperature distribution and achieves the lowest RMSE among the compared methods (0.74 K).

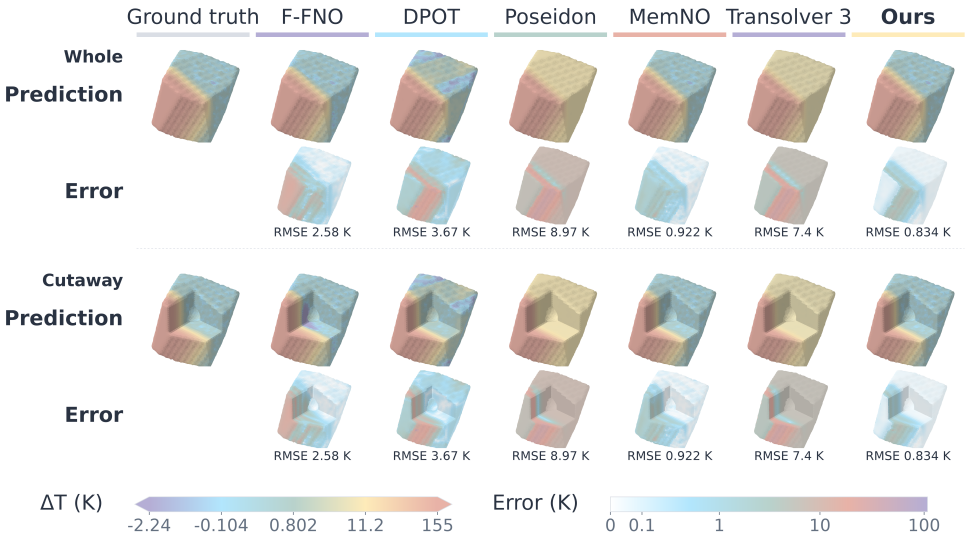

Figure 13: **Qualitative comparison for 3D Case 2.** Whole and cutaway views show ground-truth and predicted temperature variations ΔT, with absolute errors displayed below each prediction. The cutaway exposes the internal temperature distribution. Our method captures both exterior and interior temperature patterns and achieves the lowest RMSE among the compared methods (0.834 K).

C Baselines and Evaluation Details

C.1 Dataset and Observation Interface

MatTherm construction, material splits, geometry and source families, observation corruption, and trajectory generation are described in Appendix A. Standard trajectories contain 40 recorded evolution frames, whereas long-horizon trajectories contain 80; the initial field T_0 is stored separately.

Identification and few-shot adaptation use perturbed temperature observations. Temperature-prediction metrics are evaluated against the corresponding clean trajectories. The reference conductivity and source remain available for supervision and evaluation diagnostics. Under the original operating condition, UniMaNO does not use the reference source at inference and instead predicts the heat source jointly with conductivity. For transfer to a new operating condition, the recovered conductivity is combined with the prescribed source of the target condition.

No temperature frame belonging to an evaluated prediction interval is used as an input to the corresponding prediction.

C.2 Experimental Settings

Long-horizon prediction. Table 1 evaluates 500 seen materials per dimension, with 2D and 3D models trained separately. Frames 1 through 40 form the observation window, and frames 41 through 80 form the prediction window.

UniMaNO uses frames 1 through 40 to infer the conductivity function and the heat source associated with the observed operating condition. Starting from the temperature field at frame 40, the PDE solver advances the system through frames 41 through 80 using the inferred conductivity and model-predicted source.

The forward neural-operator baselines use their native 40-frame temporal context. They receive frames 1 through 40 and recursively predict frames 41 through 80, inserting each predicted field into the rolling context before the next step. Every method is evaluated on the same 40 future frames.

Cross-condition prediction. Table 2 evaluates 500 unseen materials per setting in 2D. Each material is associated with an identification experiment A and a target experiment B. The material law is shared between the two experiments, while the heat source, geometry, or both may change.

UniMaNO infers conductivity from frames 1 through 40 of A. The inferred conductivity is then held fixed and reused in B together with the target geometry, prescribed heat source, and initial temperature field. The PDE solver generates the target trajectory directly from these quantities.

Forward neural operators operate on the target experiment using their native 40-frame context. They receive frames 1 through 40 of B and recursively predict frames 41 through 80. All methods are evaluated against the clean target trajectory over frames 41 through 80.

UniMaNO-ZS performs direct identification from A. UniMaNO-FS uses the same complete identification sequence and takes its first five recorded frames as reconstruction supervision for adaptation. These frames are already contained in the identification observations.

Protocol comparability. Long-horizon prediction uses the same observation and evaluation windows for all methods. Cross-condition prediction follows the native inference interface of each method: UniMaNO transfers a material law identified from A, while forward operators use the target history from B. The comparison shares the same target prediction interval and clean evaluation trajectory, and no method receives temperatures from frames 41 through 80 during prediction.

Conductivity identification. Table 3 evaluates three seen and three unseen materials. Inverse PINN, gPINN, and mODIL jointly recover conductivity and the unknown identification-stage heat source through per-instance optimization. UniMaNO-ZS directly predicts conductivity from the observed sequence, while UniMaNO-FS additionally performs five-frame reconstruction-based refinement. Full L2 and Obs. L2 evaluate the recovered conductivity over the complete modeled temperature range and the range covered by the identification observations, respectively. The PDE column evaluates transfer trajectories using the reference conductivity.

Representation analysis and ablations. Figure 4 visualizes the learned material space using t-SNE. Selected representations and their midpoints are decoded and simulated under identical physical conditions. Figure 5 compares inferred, mean, and random representations for an unseen case.

For Figure 6, simulation supervision and same-material consistency are removed independently while all remaining settings are fixed. All variants use zero-shot evaluation. Tail RMSE covers the final quarter of the long-horizon prediction interval, corresponding to frames 71 through 80, and Energy covers the complete prediction interval.

C.3 Baseline Configurations

Neural prediction baselines. F-FNO (Tran et al., 2023), DPOT (Hao et al., 2024), Poseidon (Herde et al., 2024), MemNO (Buitrago Ruiz et al., 2025), and Transolver3 (Zhou et al., 2026) are trained on MatTherm using the architectures in Table 18. All models use randomly initialized task-specific parameters.

Each model predicts the next temperature field from a 40-frame temperature context together with the source inputs, geometry information, spatial coordinates, and observation masks used by the corresponding forward model. The 2D geometry input contains the domain mask; the 3D input additionally contains cell-volume fractions and signed distances.

Temperatures are centered at 300 K and scaled by 100 K. Source fields are scaled by 700, and predicted temperature increments use a scale of 20 K. Temperature features contain absolute temperature, change relative to the first context frame, and successive temporal differences. During autoregressive prediction, the newest prediction enters the 40-frame context for the subsequent step.

Training follows the same 40-frame context interface used at inference. All baselines use AdamW with weight decay 10^{-6}, with batch sizes 8 in 2D and 1 in 3D. AdamW uses $(0.9, 0.999)$ for its momentum parameters, except DPOT, which uses $(0.9, 0.9)$. Model-specific learning rates, gradient clipping, and schedules are reported in Table 19. Checkpoints are selected using autoregressive validation response error.

Table 18: Neural baseline architectures used in the experiments. Width and depth describe the spatial backbone; modes denote spectral-mode counts.

Method	2D	3D
F-FNO	Width 64; depth 12; modes 32	Width 64; depth 8; modes 12
DPOT	Width 512; depth 4; modes 32	Width 256; depth 4; modes 8
Poseidon	Width 128; depth 8; 8 heads	Width 96; depth 6; 8 heads
MemNO	Width 128; depth 8; modes 24; memory 128	Width 72; depth 6; modes 12; memory 72
Transolver3	Width 256; depth 8; 8 heads; 64 slices	Width 256; depth 6; 8 heads; 48 slices

Table 19: Optimization settings for the neural prediction baselines.

Method	Learning rate	Gradient clip	Schedule
F-FNO	10^{-3}	1	One-cycle; warm-up fraction 0.1; initial division factor 25
DPOT	10^{-3}	5	One-cycle; warm-up fraction 0.2; initial division factor 10^4
Poseidon	5×10^{-4}	5	Cosine decay
MemNO	10^{-3}	1	45,000-update linear warm-up followed by cosine decay
Transolver3	10^{-3}	1	Cosine decay

Inverse PINN and gPINN. Inverse PINN (Raissi et al., 2019) jointly fits temperature, conductivity, and nonnegative source networks. The temperature and conductivity networks use hidden width 64 with $\tanh$ activations. Conductivity is constrained to $[10^{-3}, 256]$. The source uses a rank-eight separable representation with softplus outputs.

The objective combines observation, initial-condition, boundary-condition, PDE, conductivity-smoothness, and source-regularization terms. gPINN (Yu et al., 2022) additionally penalizes gradients

of the normalized PDE residual with respect to normalized space-time coordinates. Both methods use 5,000 Adam updates with learning rate 10^{-3} and gradient-norm clipping at 10.

mODIL. mODIL (Karnakov et al., 2023) jointly optimizes a multilevel spatial representation of the temperature trajectory together with conductivity and source networks. Its discrete PDE loss uses harmonic face diffusivities, no-flux boundaries, and implicit-midpoint integration with two substeps per recorded interval. The initial field is imposed directly. Optimization uses 5,000 Adam updates with learning rate 10^{-2} and gradient-norm clipping at 10. All per-instance inversion baselines report their final optimization iterate.

C.4 Evaluation Metrics and Hardware

Temperature errors. Let Ω_h denote the active spatial cells, $\mathcal{P}$ the evaluated prediction frames, and T^c the clean reference trajectory. We define

$$\|V\|^2 = \sum_{n\in\mathcal{P}} \sum_{j\in\Omega_h} V_{n,j}^2. \tag{9}$$

The absolute prediction error is

$$\mathrm{RMSE} = \sqrt{\frac{\|\widehat{T} - T^c\|^2}{|\mathcal{P}|\,|\Omega_h|}}. \tag{10}$$

The response-normalized error is

$$\mathrm{L2} = \sqrt{\frac{\|\widehat{T} - T^c\|^2}{\max(\|T^c - T_0\|^2, \varepsilon)}}. \tag{11}$$

RMSE is reported in kelvin. In the L2 denominator, the initial field of the evaluated trajectory is broadcast over the prediction interval. For transfer prediction, T_0 belongs to the target experiment. Tail RMSE uses the final quarter of the corresponding prediction interval.

PDE residual. For each evaluated frame, the reference solver advances the preceding predicted state by one recorded interval using the reference conductivity and reference source:

$$T_n^{\mathrm{phys}} = \mathcal{S}_{\Delta t,g}(\widehat{T}_{n-1}; k^\star, q_n^\star). \tag{12}$$

The predicted recorded-step rate is

$$A_n = \frac{\widehat{T}_n - \widehat{T}_{n-1}}{\Delta t}. \tag{13}$$

The reference-dynamics rate is

$$B_n = \frac{T_n^{\mathrm{phys}} - \widehat{T}_{n-1}}{\Delta t}. \tag{14}$$

The symmetric relative residual is

$$\mathrm{PDE} = \frac{2\|A - B\|}{\max(\|A\| + \|B\|, \varepsilon)}. \tag{15}$$

For the first evaluated frame, the preceding state is the state that initializes the prediction interval. The same definition is used for the transfer-trajectory PDE values reported in Table 3.

Energy balance. With no-flux boundaries, conductive face fluxes cancel after spatial summation. Following the recorded-step convention used throughout the evaluation, the total predicted temperature-change rate is

$$U_n = \sum_{j\in\Omega_h} A_{n,j}. \tag{16}$$

The corresponding reference net-heating rate is

$$V_n = \sum_{j \in \Omega_h} \left[q^\star_{n,j} - \beta(\widehat{T}_{n-1,j} - T_{\mathrm{amb}}) \right]. \tag{17}$$

The reported energy-balance error is

$$\mathrm{Energy} = \frac{1}{|\mathcal{P}|} \sum_{n \in \mathcal{P}} \frac{2|U_n - V_n|}{\max(|U_n| + |V_n|, \varepsilon)}. \tag{18}$$

This diagnostic is defined at recorded-frame resolution. Internal numerical substeps affect trajectory integration only and do not alter the metric definition. The uniform cell-volume factor cancels in the normalized ratio. PDE and Energy are dimensionless, and ε is set to 10^{-12}.

Conductivity errors. For a temperature query grid Ξ, conductivity recovery is measured by

$$\mathrm{RelL2}_k(\Xi) = \frac{\sqrt{\sum_j [\widehat{k}(\xi_j) - k^\star(\xi_j)]^2}}{\max\left(\sqrt{\sum_j k^\star(\xi_j)^2}, 10^{-12}\right)}. \tag{19}$$

Full L2 uses the 256-point grid over $[200, 1000]$ K. For Obs. L2, we take the minimum and maximum temperatures covered by valid identification observations together with the active cells of T_0, intersect this interval with $[200, 1000]$ K, and place 256 uniformly spaced query temperatures between its endpoints. Predicted and reference conductivity functions are linearly interpolated at these temperatures.

Aggregation and identification time. Metrics are computed per scene, averaged within each material, and then averaged equally across materials. Identification time is measured per instance with GPU synchronization. Zero-shot timing includes network inference and conductivity decoding. Few-shot timing additionally includes adaptation. Per-instance inversion timing includes initialization, optimization, and final conductivity decoding. Data loading, checkpoint loading, diagnostic output, and downstream thermal simulation are excluded.

Hardware. Experiments use PyTorch with CUDA on NVIDIA GeForce RTX 5090 GPUs with 32 GB memory. Identification timing is measured on the same hardware.

D Implementation Details

D.1 Representation and Network Architecture

Stage I representation. Stage I applies POD to centered log-conductivity curves evaluated on the shared 256-point temperature grid and retains R of 32 modes. Projection coefficients are standardized using their training-set means and standard deviations, with a standard-deviation floor of 10^{-6}. The resulting shift and scale are absorbed into the fixed trunk representation described in Section 3.2. The shared basis, normalization statistics, and reference material representations remain fixed throughout Stage II.

Input features. The 2D observation encoder uses five channels: $(T_{\mathrm{obs}} - 300)/100$, $(T_{\mathrm{obs}} - T_0)/100$, successive temperature differences divided by 100, the domain mask, and the observation mask. The first temporal difference is measured relative to T_0, while later differences require valid observations at both adjacent times.

The 3D encoder additionally uses cell-volume fractions and signed distances, giving seven channels. Signed distances are divided by 0.25 and clipped to $[-1, 1]$. Temperature features at missing locations are masked.

Encoder and prediction heads. The encoder contains four spatial levels with channels $(24, 48, 96, 128)$ in 2D and $(12, 24, 48, 64)$ in 3D. Each block combines spatial and temporal convolutions, residual connections, GroupNorm, and SiLU. Spatial resolution is reduced after the first level while temporal resolution is preserved.

A U-Net-style decoder predicts the nonnegative heat-source field:

$$\widehat{q} = 700 \ \mathrm{softplus}(q_{\mathrm{raw}}). \tag{20}$$

The output bias is initialized to -4.

The material branch combines encoder features with source intensity, log-transformed source intensity, temporal source differences, and domain coverage. Spatial pooling and temporal attention aggregate these features across the observation sequence before fusion with the geometry encoding. The branch outputs 32 material coefficients. Feature and hidden dimensions are $128/128$ in 2D and $64/96$ in 3D. The coefficient head uses LayerNorm, SiLU, and dropout with probability 0.1.

D.2 Differentiable Simulator

The differentiable simulator uses conservative flux differences on active Cartesian cells. At each numerical substep, the diffusivity at cell j is

$$D_j = a\widehat{k}(T_j). \tag{21}$$

For adjacent active cells j and l, the face diffusivity is obtained by harmonic averaging:

$$D_{jl} = \frac{2D_j D_l}{D_j + D_l + 10^{-12}}. \tag{22}$$

Normal diffusive flux is zero on external boundaries and internal holes or cavities. Explicit Euler integration uses a substep size satisfying

$$\delta t \le \frac{0.9}{2D_{\max} \sum_d h_d^{-2}}, \tag{23}$$

where h_d is the grid spacing along spatial dimension d.

The reported training runs use at most 16 Euler substeps within each recorded interval. The source is held constant within a recorded interval, while conductivity and diffusive fluxes are reevaluated from the evolving temperature at every substep. Gradients propagate through both conductivity evaluation and temperature updates.

The conductivity representation is defined over $[200, 1000]$ K. MatTherm training trajectories remain within the $[200, 600]$ K thermal training range specified in Appendix A, so training remains inside the support of the conductivity representation. The support interval defines where conductivity is represented and queried; evaluation metrics are computed directly from the recorded temperature fields.

Internal substeps are part of trajectory integration. All reported temperature, PDE, and Energy metrics are evaluated at the recorded-frame resolution defined in Appendix C.

D.3 Training Losses and Optimization

Simulation reconstruction. For scene i, let $m_{n,j}^{(i)}$ indicate a valid observation inside the material domain. The normalized simulation loss is

$$\mathcal{L}_{\mathrm{sim}}^{(i)} = \frac{\sum_{n,j} m_{n,j}^{(i)} (\widehat{T}_{n,j}^{(i)} - T_{\mathrm{obs},n,j}^{(i)})^2}{\max\Big(\sum_{n,j} m_{n,j}^{(i)} (T_{\mathrm{obs},n,j}^{(i)} - T_{0,j}^{(i)})^2, 10^{-8}\Big)}. \tag{24}$$

Let s_r denote the Stage I standard deviation of coefficient r. Its normalized importance weight is

$$w_r = \frac{s_r^2}{\sum_l s_l^2}. \tag{25}$$

Material-representation supervision is

$$\mathcal{L}_z^{(i)} = \sum_{r=1}^{R} w_r(\widehat{z}_{ir} - z_{ir}^{\star})^2. \tag{26}$$

For same-material groups G within a batch, representation consistency is

$$\mathcal{L}_{\text{inv}} = \frac{1}{|\mathcal{B}|} \sum_{G\in\mathcal{B}} \frac{1}{|G|} \sum_{i\in G} \sum_{r=1}^{R} w_r(\widehat{z}_{ir} - \overline{z}_{G,r})^2. \tag{27}$$

Each Stage II batch contains two scenes from the same material.

Source supervision. The source loss combines weighted absolute error with a spatial-difference penalty:

$$\mathcal{L}_q^{(i)} = \frac{\sum_{n,j} \omega_{n,j}^{(i)} |\widehat{q}_{n,j}^{(i)} - q_{n,j}^{\star(i)}|}{700 \sum_{n,j} \omega_{n,j}^{(i)}} + 0.1\mathcal{L}_{\nabla q}^{(i)}. \tag{28}$$

The spatial weight is

$$\omega_{n,j}^{(i)} = \mathbf{1}_{\Omega_h^{(i)}}(j) \left[1 + 4\mathbf{1}\{q_{n,j}^{\star(i)} > 7\}\right]. \tag{29}$$

The spatial-difference term averages the absolute error of source differences between neighboring active cells over space and time, then averages across spatial directions and normalizes by 700.

Training schedule. Stage II initially conditions the material branch on the reference source, then gradually replaces this conditioning with the model prediction. The conditioning field is

$$q_{\text{cond}} = (1 - \rho_e)q^{\star} + \rho_e \widehat{q}. \tag{30}$$

During epochs 1 through 10, ρ_e is zero. During epochs 11 through 30, it increases linearly from zero to one. From epoch 31 onward, the material branch uses the predicted source and simulation supervision is enabled.

The source-loss weight remains 1 and the same-material consistency weight remains 0.05. The material-representation weight starts at 0.25 and decreases to 0.025, reaching its final value at epoch 101 in 2D and epoch 81 in 3D.

Optimization. Stage II uses AdamW with learning rate 3×10^{-4}, weight decay 10^{-6}, batch size 2, and gradient-norm clipping at 1 for at most 1,000 epochs. Validation is performed every five epochs.

Checkpoint selection combines the mean simulation reconstruction loss with relative source error and material-representation error. The source term is weighted by 0.25, and the representation term is weighted by 0.1.

Early stopping is enabled at epoch 105 in 2D and epoch 85 in 3D, with patience of six validation evaluations. After two evaluations without sufficient improvement, the learning rate is halved down to 10^{-6}. Supported network operations use bfloat16, while material-feature reductions and differentiable simulation use float32. The 3D model additionally uses activation checkpointing in the encoder and simulator.

D.4 Zero-Shot and Few-Shot Inference

Zero-shot identification. UniMaNO-ZS predicts the material representation and identification-stage heat source in a single forward pass. The material representation is decoded with the fixed Stage I trunk, and network parameters remain unchanged.

For long-horizon prediction, conductivity and heat source are inferred from frames 1 through 40 of the observed trajectory. Starting from frame 40, the PDE solver advances the system through frames 41 through 80 using these inferred physical quantities.

For cross-condition prediction, conductivity is inferred from experiment A and held fixed. Experiment B is then simulated using the recovered conductivity together with the target geometry, prescribed heat source, and initial temperature field.

Five-frame adaptation. UniMaNO-FS refines identification through simulation-based reconstruction using observations already contained in the identification sequence. The complete observation sequence remains the model input, while the first five recorded frames provide adaptation supervision.

The spatiotemporal encoder, source decoder, geometry encoder, and material head are initialized from the pretrained parameters θ_0 and adapted jointly. The Stage I basis and differentiable simulator remain fixed.

Adaptation follows equation 8 using equation 24 restricted to the five supervised frames, with a parameter-prior weight of 10^{-4}. We perform 20 AdamW updates with learning rate 10^{-5}, zero weight decay, and gradient-norm clipping at 1. Dropout is disabled during adaptation. Each scene starts independently from the same pretrained parameters. When adaptation increases the reconstruction error on the supervised frames, the zero-shot identification is retained.

For cross-condition prediction, both UniMaNO-ZS and UniMaNO-FS identify the material from experiment A. Few-shot adaptation uses five frames already contained in A. Neither mode receives temperature observations from the evaluated prediction interval of B.

Prediction interface. Under the original operating condition, thermal prediction uses the inferred conductivity and model-predicted heat source. Under a new operating condition, the inferred conductivity is reused with the target geometry, prescribed source, and initial temperature field. Forward neural operators propagate temperature directly through their native 40-frame autoregressive context.