CROSS-PHYSICS MAPPING: LEARNING OPERATORS BETWEEN FUNDAMENTALLY DIFFERENT PHYSICAL DYNAMICS

Anonymous authors
Paper under double-blind review

ABSTRACT

Neural operators have demonstrated remarkable success in learning solution mappings within individual physical systems, yet whether they can learn transformations between fields governed by fundamentally different partial differential equations remains largely unexplored. We introduce **Cross-Physics Mapping (CPM)**, an operator-learning framework for translating heterogeneous physical fields through compatible latent representations without assuming equivalence between their governing dynamics. We formulate conditions under which cross-physics operators are well-defined and introduce a dimensionless scaling principle for constructing physically compatible training pairs. To investigate this problem, we generate 1,000 paired diffusion and wave realizations with shared latent geometries, material heterogeneities, and excitation configurations, while restricting network inputs to surface-accessible spatiotemporal observations. We evaluate seven reported model configurations, including six neural operators and a convolutional baseline, in both mapping directions. The results reveal a consistent directional asymmetry on this dataset: diffusion-to-wave mapping is more difficult because diffusion attenuates high-frequency and phase-sensitive information required for reconstructing propagating wavefronts. The U-shaped Neural Operator obtains a relative ℓ_2 error of 0.307 and an R^2 of 0.905 in this direction, whereas the Galerkin Neural Operator obtains 0.154 and 0.935, respectively, for wave-to-diffusion mapping. Because one diffusion-to-wave FNO configuration uses additional gradient supervision, the comparison is a study of complete model–training configurations rather than a controlled architectural ranking. The findings establish the feasibility of conditional cross-physics prediction under shared latent structure while delimiting it by source information, training distribution, and physical admissibility.

1 INTRODUCTION

Can a model translate observations governed by one PDE into fields governed by a fundamentally different PDE? This problem differs from standard surrogate modeling: the input is already the response of one physical system, not the initial condition or coefficient field of the target system. For example, diffusion suppresses fine spatial structure, whereas wave propagation retains oscillatory phase and time-of-flight information. A useful mapping must therefore exploit information shared by the two measurements without assuming that their dynamics are equivalent.

The distinction is consequential for both learning and interpretation. A low reconstruction error does not imply that the source observation uniquely determines the target: a network can exploit correlations in the training distribution when the physical inverse is ambiguous. Cross-physics learning therefore requires an explicit account of identifiability, information loss, and the construction of compatible pairs. Neural operators learn mappings between function spaces and have been successful for parametric PDEs Lu et al. (2019); Li et al. (2020); Kovachki et al. (2023b); Azizzadenesheli et al. (2024). Cross-modal sensing poses a different question: whether two physical observations share a latent representation that is sufficiently informative for conditional prediction. This question is relevant to nondestructive evaluation, where thermal, acoustic, optical, and electromagnetic measurements encode the same structure through different transfer mechanisms.

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.

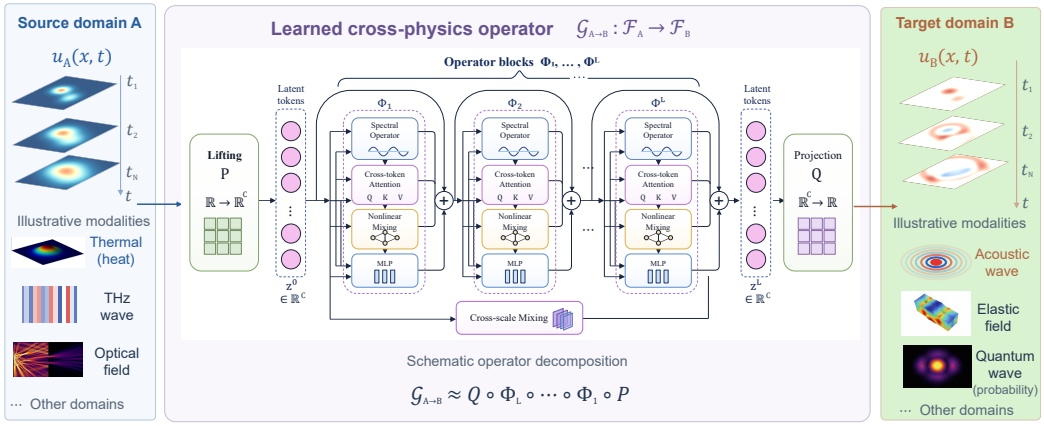

Figure 1: Conceptual illustration of Cross-Physics Mapping (CPM). A physical field $u_A(x,t)$ from a source domain is mapped to a field $u_B(x,t)$ in a different target domain through a learned operator $\mathcal{G}_{A\to B}$. The source and target domains may correspond to distinct physical phenomena, including thermal, optical, acoustic, elastic, or quantum systems. The mapping is formulated at the function-space level and can be approximated by different learning architectures.

We introduce *Cross-Physics Mapping* (CPM) (see Figure 1), which learns

$$\mathcal{G}_{A\to B} : \mathcal{F}_A \to \mathcal{F}_B, \qquad \widehat{u}_B = \mathcal{G}_{A\to B}(u_A), \tag{1}$$

from paired observations associated with a shared latent physical configuration. We study diffusion and wave propagation, generate each field independently from its own PDE, and align their characteristic spatial reach through dimensionless scaling. Seven model configurations are evaluated in both directions. The results show a consistent asymmetry: diffusion-to-wave (D2W) prediction is harder because the source field has attenuated information needed to reconstruct propagating structures. Our contributions are:

- a latent-variable formulation with explicit existence and stability conditions for mappings between observations governed by different PDEs;
- a dimensionless alignment rule that pairs characteristic evolution scales without claiming dynamical equivalence; and
- a bidirectional diffusion–wave benchmark across seven model configurations, exposing the asymmetry created by dissipative information loss.

D2W predictions are consequently interpreted as conditional estimates under the training distribution, not unconditional recovery of information erased by diffusion.

2 Related Work

Scientific machine learning and neural operators. Physics-informed neural networks (PINNs) incorporate governing equations and boundary or initial conditions into the training objective Raissi et al. (2018a); Karniadakis et al. (2021), while operator-learning methods aim to approximate mappings between function spaces rather than individual PDE solutions. DeepONet and the Fourier neural operator (FNO) established two widely used formulations of this idea Lu et al. (2019); Li et al. (2020), followed by theoretical and methodological developments clarifying approximation properties, discretization invariance, and the broader neural-operator framework Berner et al. (2026). Physics-informed neural operators further combine operator learning with PDE constraints, allowing physical residuals to complement data supervision across resolutions. Subsequent architectures have introduced multiresolution Fourier processing, wavelet representations, compressed latent spaces, and attention-based global interaction Rahman et al. (2022); Tripura & Chakraborty (2023); Wang & Wang (2024). Graph- and geometry-aware operators extend these ideas to irregular meshes and

non-Cartesian domains Li et al. (2023b); Kovachki et al. (2023a); Wen et al. (2026), while GNOT and FactFormer use transformer-style architectures for heterogeneous inputs and scalable PDE surrogate modeling Hao et al. (2023); Li et al. (2023a). Our experiments cover representative convolutional, spectral, wavelet, latent, and Galerkin operator families, but not the complete neural-operator or transformer literature; the comparison should therefore be interpreted as a study of representative model configurations rather than a comprehensive state-of-the-art ranking.

Learning across physical systems. Most neural-operator studies learn a solution map within a fixed governing equation or a parameterized family of closely related PDEs, for example from coefficients, initial conditions, forcing terms, or geometry to the corresponding solution field Li et al. (2020); Kovachki et al. (2023a); Raissi et al. (2018b); Goswami et al. (2022); Kurth et al. (2023); You et al. (2022). Recent work has begun to investigate broader transfer across PDE families, including pretrained or foundation-style models whose representations can be adapted to previously unseen dynamics Sun et al. (2025); Herde et al. (2024); Kim et al. (2026). These studies primarily transfer a model across governing equations or tasks. CPM considers a different setting: both the input and target are already physical response fields, generated independently by distinct PDEs, and the objective is to learn a conditional operator between those observations through their shared latent configuration. Thus, the problem is closer to cross-domain field translation than to conventional parameter-to-solution surrogate modeling.

Multimodal sensing and cross-physics prediction. Different sensing modalities may probe the same geometry or material state while encoding substantially different spatial, temporal, and spectral information. Multimodal learning has therefore been used to combine or translate heterogeneous observations in scientific imaging, remote sensing, medical imaging, and nondestructive evaluation Zhu et al. (2026d); Zhu & Maldague (2026). However, generic image-to-image translation does not normally distinguish whether the two fields satisfy different evolution equations, nor does it explicitly address whether one physical observation contains sufficient information to determine another. A more closely related line of work arises in inverse and virtual-field reconstruction. Analytical virtual-wave methods relate diffusion measurements to an auxiliary wave field through an ill-conditioned temporal transform Burgholzer et al. (2017); Zhu et al. (2026c;a;b). Such constructions exploit a known analytical relationship between the representations and require regularization because diffusion suppresses high-frequency information. Related inverse-problem literature has likewise emphasized the instability associated with recovering fine-scale or phase-sensitive information from smoothing forward operators Beck (1970); Wang & Zabaras (2004); Beck et al. (1996). CPM differs in that its target is not an analytically constructed virtual field: the diffusion and wave observations are generated independently from their own PDEs and are coupled only through a common latent scene. Consequently, external virtual-wave reconstruction errors are not directly comparable to CPM errors; the conditions required for a controlled comparison are described in Appendix 5.

Benchmarks and general-purpose PDE learning. Several benchmarks have been introduced to make evaluation in scientific machine learning more systematic. PDEBench covers multiple canonical PDE systems and standardizes data generation and surrogate evaluation Takamoto et al. (2022). More recent large-scale datasets and pretrained models broaden the range of physical systems considered Subramanian et al. (2023); Alkin et al. (2024); McCabe et al. (2024), but these benchmarks predominantly formulate tasks as evolution prediction or parameter-to-solution learning within each physical system. Their samples do not, in their standard form, provide paired responses of fundamentally different PDEs generated from the same latent geometry, material heterogeneity, and excitation configuration. PDEBench and related benchmarks are therefore useful for testing generic surrogate capacity, whereas the present dataset is designed to isolate cross-PDE factorization, conditional identifiability, and directional information loss.

3 METHOD

3.1 CROSS-PHYSICS FACTORIZATION

Let $z \in \mathcal{Z}$ encode the shared geometry, material heterogeneity, excitation, and characteristic scales. With solution operators $\mathcal{S}_A, \mathcal{S}_B$ and observation operators $\mathcal{O}_A, \mathcal{O}_B$, the paired fields are

$$u_A = \mathcal{A}(z) = \mathcal{O}_A \mathcal{S}_A(z), \qquad u_B = \mathcal{B}(z) = \mathcal{O}_B \mathcal{S}_B(z). \tag{2}$$

The PDEs and material parameters may differ; only the latent realization is paired. Let $\mathcal{A} : \mathcal{Z} \to \mathcal{F}_A$ and $\mathcal{B} : \mathcal{Z} \to \mathcal{F}_B$ be two maps. A deterministic map $\mathcal{G}_{A\to B}$ satisfying $\mathcal{B} = \mathcal{G}_{A\to B} \circ \mathcal{A}$ exists if and only if

$$\mathcal{A}(z_1) = \mathcal{A}(z_2) \implies \mathcal{B}(z_1) = \mathcal{B}(z_2). \tag{3}$$

If, in addition, $\mathcal{Z}$ is a compact metric space, $\mathcal{F}_A$ and $\mathcal{F}_B$ are normed spaces, and $\mathcal{A}$ and $\mathcal{B}$ are continuous, then the induced map is continuous on $\mathcal{A}(\mathcal{Z})$.

3.2 Dimensionless Alignment

We independently solve heterogeneous diffusion and wave equations,

$$\partial_t T = \nabla\cdot[\alpha(\mathbf{x})\nabla T] + s_D, \qquad \partial_{tt} p = \nabla\cdot[c^2(\mathbf{x})\nabla p] + s_W. \tag{4}$$

For characteristic length L, observation time t_0, background diffusivity α_b, and wave speed c_b, nondimensionalization yields

$$\mathrm{Fo} = \frac{\alpha_b t_0}{L^2}, \qquad \Lambda = \frac{c_b t_0}{L}. \tag{5}$$

The diffusion length is $\ell_D \sim \sqrt{\alpha_b t_0}$ and the wave travel distance is $\ell_W = c_b t_0$. We match normalized spatial reach by imposing

$$\sqrt{\mathrm{Fo}} = \Lambda \quad\Longleftrightarrow\quad c_b = \sqrt{\alpha_b / t_0}. \tag{6}$$

This is a data-construction rule, not a physical velocity for Fourier conduction or an equivalence between parabolic and hyperbolic dynamics. Full derivations and the heterogeneous extension are in Appendix B.

3.3 Paired Data and Learning

Each latent realization specifies defect geometry, excitation, and physical coefficients. For defect j, let (x'_j, y'_j) denote coordinates rotated by ψ_j. Elliptical inclusions are represented by

$$r^{(2)}_{d,j}(x,y) = \left(\frac{x'_j}{a_j}\right)^2 + \left(\frac{y'_j}{b_j}\right)^2, \tag{7}$$

whereas rectangular inclusions use

$$r^{(\infty)}_{d,j}(x,y) = \max\left(\frac{|x'_j|}{a_j}, \frac{|y'_j|}{b_j}\right). \tag{8}$$

A square is obtained by setting $a_j = b_j$. In either case, the corresponding smooth mask is

$$m_j(x,y) = \frac{1}{2}\left[1 - \tanh\left(\frac{r_{d,j}(x,y) - 1}{\varepsilon_d}\right)\right]. \tag{9}$$

For multiple defects, the union mask is

$$m(x,y) = 1 - \prod_{j=1}^{N_d}\left[1 - m_j(x,y)\right]. \tag{10}$$

The geometry is shared, but the contrasts, sources, units, and governing operators remain modality-specific. The two PDEs are then solved independently; the target is never obtained by transforming the source field. Only the same accessible boundary Γ_s is retained:

$$u_D = \mathcal{S}_D(z)|_{\Gamma_s}, \qquad u_W = \mathcal{S}_W(z)|_{\Gamma_s}. \tag{11}$$

The dataset contains 1,000 paired 501×200 spatiotemporal fields. We split it by realization into 800 training, 100 validation, and 100 test samples. Each modality is normalized using training-set statistics. Simulation, source, boundary, and sampling details are reported in Appendix C. We train separate D2W and wave-to-diffusion (W2D) models. To examine whether cross-physics mapping depends on a particular representation strategy, we evaluate seven architectures spanning convolutional, spectral, multiresolution, latent-space, basis-function, and attention-based operator learning.

All models receive the same surface spatiotemporal observation as input and predict the corresponding target field on the same observation grid. Detailed architectures and implementation parameters are provided in Appendix D.

ResUNet serves as a convolutional baseline. It employs a residual encoder–decoder with multiscale skip connections, allowing local and hierarchical spatiotemporal features to be transferred between the source and target physical fields. Unlike the neural operators considered below, ResUNet does not explicitly parameterize a function-space operator. **DeepONet** represents the cross-physics transformation through separate branch and trunk networks. The branch network encodes the complete source field, whereas the trunk network represents the target coordinates (t, x). Their contraction produces the predicted field, providing a basis-function representation of the learned operator. **Fourier Neural Operator (FNO)** learns nonlocal field interactions in the spectral domain. Fourier modes of the spatiotemporal representation are transformed by learned spectral kernels and combined with local pointwise mappings, enabling global dependencies to be modeled directly at the original observation resolution. **U-Shaped Neural Operator (U-NO)** combines Fourier operator blocks with a U-shaped multiresolution hierarchy. Progressive compression enlarges the effective receptive field, while skip connections preserve information across spatiotemporal scales. This construction is particularly suited to mappings that contain both large-scale structures and localized high-frequency features. **Wavelet Neural Operator (WNO)** replaces the global Fourier representation with a multiresolution wavelet decomposition. Learnable transformations are applied to different wavelet subbands, allowing smooth and rapidly varying field components to be represented separately across scales. **Latent Neural Operator (LNO)** first compresses the source observation into a lower-dimensional latent field and performs operator learning within this compressed representation. The processed latent field is subsequently decoded to the target domain, reducing the cost of modeling global interactions while retaining a field-to-field operator structure. **Galerkin Neural Operator (GNO)** represents the field using coordinate-aware latent tokens and models their global interactions through Galerkin-type attention. In contrast to convolutional or Fourier-based representations, GNO directly mixes spatial–temporal features through a global operator-inspired attention mechanism. Together, these models provide complementary inductive biases for cross-physics learning: ResUNet emphasizes local multiscale correlations; DeepONet learns a coordinate-dependent basis expansion; FNO captures global spectral interactions; U-NO combines spectral and multiresolution representations; WNO resolves scale-dependent wavelet features; LNO performs operator learning in a compressed latent space; and GNO models global interactions through Galerkin attention. The comparison is therefore intended to test the robustness of CPM across different operator representations rather than to establish a universal ranking of neural architectures.

For prediction $\widehat{\mathbf{U}}$ and target $\mathbf{U}$, the common objective is

$$\mathcal{L}_{\mathrm{rec}} = \frac{1}{N_g}\|\widehat{\mathbf{U}} - \mathbf{U}\|_2^2 + 0.1\frac{\|\widehat{\mathbf{U}} - \mathbf{U}\|_2}{\|\mathbf{U}\|_2 + \epsilon}. \tag{12}$$

Models are optimized with AdamW, cosine learning-rate annealing, and validation-based checkpoint selection. The reported D2W FNO configuration uses $\mathcal{L}_{\mathrm{FNO}} = \mathcal{L}_{\mathrm{rec}} + 0.02\mathcal{L}_t + 0.02\mathcal{L}_x$, where $\mathcal{L}_t$ and $\mathcal{L}_x$ compare temporal and spatial gradients. We therefore compare complete model–objective configurations, not architectures in isolation. The present observation-only objective does not certify target-PDE residuals or conservation because the network does not receive complete coefficient, source, and boundary fields; physical and spectral diagnostics are discussed in Appendix A.1.

4 Experiments

4.1 Protocol and Quantitative Results

All models use the same realization-level split and are selected by validation relative ℓ_2 error. We report MSE, MAE, RMSE, relative ℓ_2, and R^2. Predictions and targets are inverse-transformed using the training-set statistics before evaluation; thus, MSE, MAE, and RMSE are reported in the original field units. Their absolute values should not be compared across mapping directions because diffusion and wave fields have different physical amplitude scales. Relative ℓ_2 is therefore the primary scale-normalized metric. Each row reports a single trained configuration under the fixed data split. Because repeated-seed runs were not conducted, close differences between models are treated as descriptive, and no statistical significance or stable architecture ranking is claimed.

The configurations were also not matched by parameter count or computational cost; the reported comparison concerns reconstruction accuracy rather than computational efficiency.

Table 1: Diffusion-to-wave reconstruction on the test set. Lower values are better for error metrics, and higher R^2 is better. †The FNO configuration uses the additional gradient terms in Appendix D; its row is not a loss-matched architecture-only comparison.

Model	MSE	MAE	RMSE	Rel. ℓ_2	R^2
ResUNet	3.03×10^{10}	4.43×10^{4}	1.74×10^{5}	0.528	0.719
DeepONet	3.25×10^{10}	4.51×10^{4}	1.80×10^{5}	0.548	0.698
FNO†	1.66×10^{10}	3.23×10^{4}	1.29×10^{5}	0.391	0.846
LNO	1.44×10^{10}	2.69×10^{4}	1.20×10^{5}	0.365	0.866
WNO	1.72×10^{10}	3.82×10^{4}	1.31×10^{5}	0.398	0.841
GNO	1.71×10^{10}	3.57×10^{4}	1.31×10^{5}	0.397	0.841
U-NO	$\mathbf{1.02 \times 10^{10}}$	$\mathbf{2.18 \times 10^{4}}$	$\mathbf{1.01 \times 10^{5}}$	**0.307**	**0.905**

Table 2: Wave-to-diffusion reconstruction on the test set. Lower values are better for error metrics, and higher R^2 is better.

Model	MSE	MAE	RMSE	Rel. ℓ_2	R^2
ResUNet	112.159	2.982	10.591	0.251	0.825
DeepONet	60.447	2.411	7.775	0.184	0.906
FNO	58.017	2.198	7.617	0.181	0.909
LNO	64.377	**1.707**	8.024	0.190	0.899
WNO	68.020	2.402	8.247	0.196	0.894
GNO	**41.903**	2.355	**6.473**	**0.154**	**0.935**
U-NO	65.644	2.234	8.102	0.192	0.898

Table 1 shows that U-NO performs best for D2W, with relative $\ell_2 = 0.307$ and $R^2 = 0.905$; LNO and the gradient-supervised FNO configuration follow at 0.365 and 0.391. Table 2 shows that GNO performs best for W2D, with relative $\ell_2 = 0.154$ and $R^2 = 0.935$, while FNO and DeepONet reach 0.181 and 0.184. Every configuration has lower relative error in W2D than in D2W. This directional consistency is more informative than a pure architecture ranking: FNO is not loss-matched in D2W, and the best model differs by direction. The training curves in Appendix E show rapid initial error reduction followed by validation plateaus. The remaining train–validation gaps are larger for several D2W configurations, consistent with the greater ambiguity of predicting oscillatory targets from smooth observations.

4.2 Representative Reconstruction

Figure 2 shows the short-pulse square-defect case. D2W targets contain propagating and defect-scattered structures absent from the smooth appearance of the source. U-NO recovers the dominant branches with the smallest global error; FNO and LNO retain the principal topology but show structured residuals, whereas ResUNet and DeepONet suppress more wave content. In W2D, several operators recover the broad thermal response from the oscillatory input. The long-pulse case tests overlapping, spatially nonuniform oscillatory structures; U-NO preserves the dominant D2W organization, while FNO and LNO retain the global response with structured fine-scale residuals. The multiple-defect case (see Figure 3) adds reflected and scattered components. U-NO recovers the dominant branches in D2W, whereas several operators reconstruct the smoother W2D target with smaller residuals. These examples support the same directional trend as the aggregate metrics without establishing phase-exact recovery. Figure 4 increases the structural complexity through multiple defects and overlapping scattered components. U-NO again retains the dominant D2W branches, while FNO and LNO preserve the global topology with more coherent residuals. The W2D targets remain smoother and are reconstructed more consistently. Defect-free chirp and localized Gaussian-chirp results are provided in Appendix E; together they show that CPM learns excitation-dependent field relations rather than only a fixed defect signature.

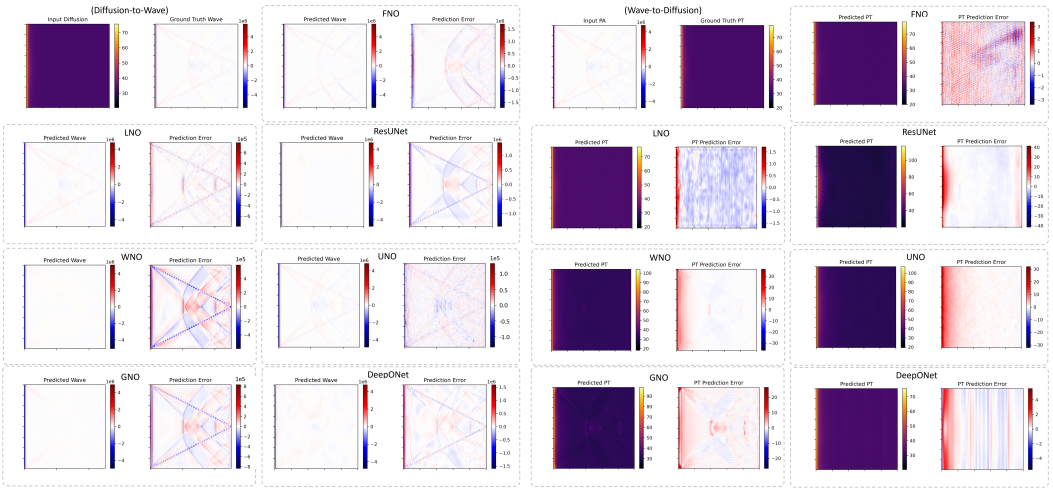

Figure 2: Bidirectional reconstruction for a square-hole defect under short-pulse excitation. Columns show the source, target, predictions, and absolute errors.

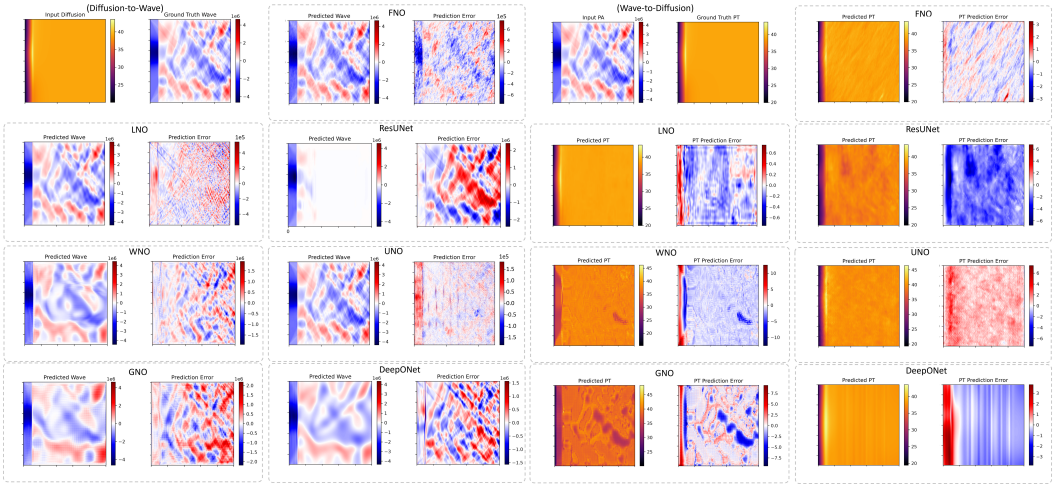

Figure 3: Bidirectional reconstruction under asymmetric Gaussian long-pulse excitation.

4.3 Directional Asymmetry

For homogeneous source-free diffusion, $\widehat{T}(\mathbf{k},t) = e^{-\alpha|\mathbf{k}|^2 t}\widehat{T}(\mathbf{k},0)$, so high-wavenumber components decay rapidly. An ideal lossless wave instead has $\widehat{p}(\mathbf{k},t) = A_{\mathbf{k}}e^{ic|\mathbf{k}|t} + B_{\mathbf{k}}e^{-ic|\mathbf{k}|t}$. D2W must infer phase-sensitive propagation features from a source in which fine scales have been attenuated; W2D maps an oscillatory observation to a smoother target. This explains the observed trend but does not prove universal identifiability or a universal ordering across all diffusion–wave systems. Global field metrics also do not certify phase fidelity; frequency-banded diagnostics are specified in Appendix A.1.

5 Mathematical Representation: Virtual-Wave Transformation

This section summarizes the analytical virtual-wave formulation developed in prior virtual-wave literature Burgholzer et al. (2017); Zhu et al. (2026c;a;b) and clarifies its relationship to the learned cross-physics mapping. The virtual-wave field is an auxiliary mathematical construction. It should

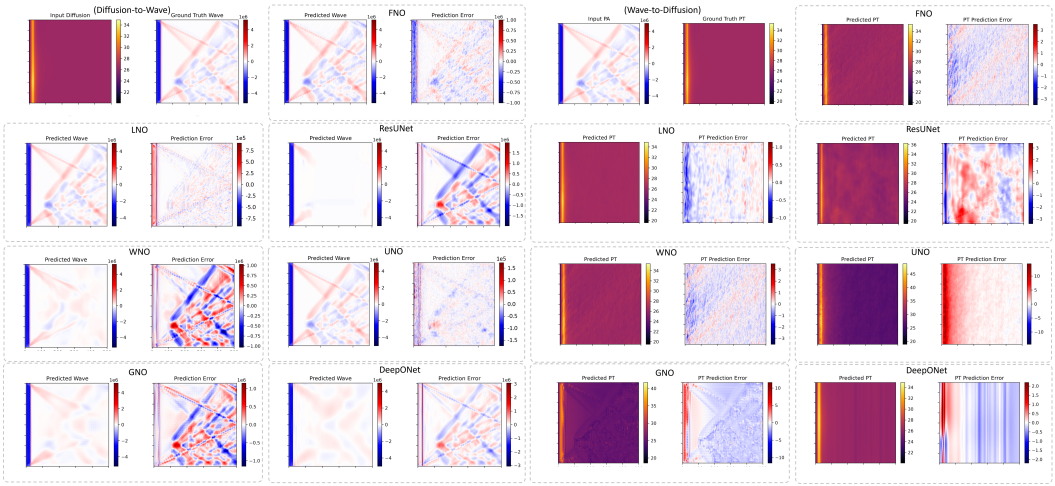

Figure 4: Bidirectional reconstruction for multiple square defects under uniform long-pulse excitation.

not be identified with a physically propagating thermal wave. For an idealized homogeneous diffusion problem with an appropriate impulsive excitation, the temperature field can be represented using an auxiliary virtual wave,

$$T(\mathbf{r},t) = \int_0^\infty K(t,t')T_{\text{virt}}(\mathbf{r},t')\,dt', \tag{13}$$

where the conventional Gaussian kernel is

$$K(t,t') = \frac{c}{\sqrt{\pi\alpha t}}\exp\left(-\frac{c^2t'^2}{4\alpha t}\right), \qquad t>0. \tag{14}$$

Here, t' denotes auxiliary virtual time and c is the selected virtual-wave velocity. The integration interval corresponds to a causal virtual-wave representation. The transformation constructs each diffusion-time observation from a weighted superposition of virtual-wave samples. Its Gaussian temporal smoothing provides the basis for the ill-conditioning of the inverse problem. After temporal discretization, the virtual-wave relation takes the form

$$\mathbf{T} = \mathbf{K}\mathbf{T}_{\text{virt}}, \tag{15}$$

where $\mathbf{K}$ is the discretized transformation kernel. Recovering the virtual-wave signal requires solving an inverse problem. A regularized formulation is

$$\mathbf{x}^* = \arg\min_{\mathbf{x}}\left[\frac{1}{2}\|\mathbf{K}\mathbf{x}-\mathbf{T}\|_2^2 + \lambda\mathcal{R}(\mathbf{x})\right], \tag{16}$$

where $\mathcal{R}$ represents a regularization functional. A regularized ADMM formulation has been used for this inverse problem Burgholzer et al. (2017); Zhu et al. (2026c)

$$\min_{\mathbf{x},\mathbf{z}} \quad \frac{1}{2}\|\mathbf{K}\mathbf{x}-\mathbf{T}\|_2^2 + \lambda\mathcal{R}(\mathbf{z}), \quad \text{subject to} \quad \mathbf{x}=\mathbf{z}. \tag{17}$$

The corresponding scaled ADMM iterations are

$$\begin{aligned}
\mathbf{x}^{k+1} &= \left(\mathbf{K}^\top\mathbf{K}+\rho\mathbf{I}\right)^{-1}\left[\mathbf{K}^\top\mathbf{T}+\rho\left(\mathbf{z}^k-\mathbf{d}^k\right)\right],\\
\mathbf{z}^{k+1} &= \text{prox}_{(\lambda/\rho)\mathcal{R}}\left(\mathbf{x}^{k+1}+\mathbf{d}^k\right),\\
\mathbf{d}^{k+1} &= \mathbf{d}^k+\mathbf{x}^{k+1}-\mathbf{z}^{k+1}.
\end{aligned} \tag{18}$$

The regularization parameter λ controls the balance between data fidelity and prior information. The ADMM penalty parameter ρ influences the numerical convergence. Regularization can stabilize the

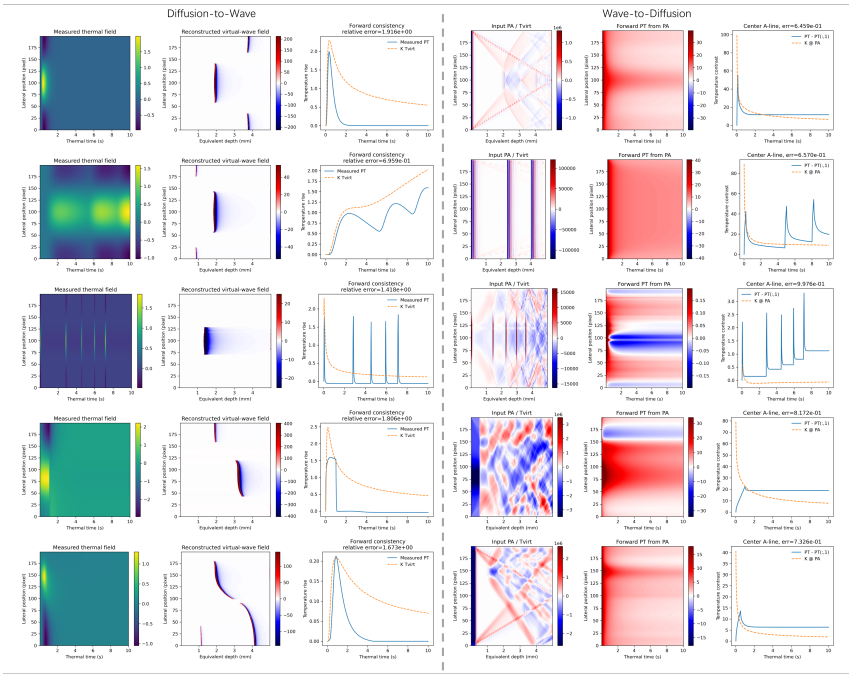

Figure 5: Representative results obtained using the conventional virtual-wave transformation under different structural and excitation conditions: a plate containing a square-hole defect under a $0.01\,\mathrm{s}$ short-pulse excitation; a defect-free plate under chirp excitation; representative cases with asymmetric, spatially nonuniform localized Gaussian long-pulse excitation of $1\,\mathrm{s}$ duration; and a plate containing multiple square defects under a uniform $1\,\mathrm{s}$ long-pulse excitation.

reconstruction, but it does not restore information that has been irreversibly attenuated by diffusion (see Figure 5). The derivation above explains why inverse temporal smoothing is ill-conditioned, but it does not by itself define a performance baseline for CPM. The analytical transform returns an auxiliary virtual-wave field, whereas the supervised target in CPM is generated independently by the wave solver. Comparing their errors is meaningful only after proving that both fields have the same source convention, coefficients, boundary conditions, observation operator, temporal support, and normalization. Accordingly, we do not import numerical errors from a different study. A controlled analytical baseline must be applied directly to all 100 held-out inputs, with one regularization-selection rule fixed without test-target access and with the same target and metrics used for the learned models. Establishing that matched experiment is outside the evidence reported in this paper. The analytical material in this appendix is retained as physical context for ill-conditioning, not as evidence of CPM superiority.

6 Conclusion

CPM formulates translation between observations governed by distinct PDEs through a shared latent configuration rather than assumed dynamical equivalence. Dimensionless length matching enables paired diffusion–wave data, while Proposition 1 separates deterministic identifiability from distribution-dependent prediction. Across seven configurations, W2D is consistently easier than D2W, reflecting the high-frequency information suppressed by diffusion. The results establish conditional cross-physics prediction on the stated synthetic distribution, not universal inversion, strict physical conservation, or a loss-controlled state-of-the-art ranking.

Acknowledgments

This work was supported by the Adolf Martens Fellowship (Grant n. BAM-AMF-2025-1).

AI USE STATEMENT

ChatGPT (OpenAI) was used to assist with manuscript drafting and language editing and to provide suggestions on mathematical formulation, methodology, and result interpretation. All AI-assisted content was independently verified, and the authors take full responsibility for the manuscript and its conclusions.

ETHICS STATEMENT

This study uses synthetic data and involves no human participants or personal information. Model predictions are estimates rather than direct physical measurements and should be independently validated before consequential engineering use.

REFERENCES

Benedikt Alkin, Andreas Fürst, Simon Schmid, Lukas Gruber, Markus Holzleitner, and Johannes Brandstetter. Universal physics transformers: A framework for efficiently scaling neural operators. *Advances in Neural Information Processing Systems*, 37:25152–25194, 2024.

Kamyar Azizzadenesheli, Nikola Kovachki, Zongyi Li, Miguel Liu-Schiaffini, Jean Kossaifi, and Anima Anandkumar. Neural operators for accelerating scientific simulations and design. *Nature Reviews Physics*, 6(5):320–328, 2024.

J. V. Beck, B. Blackwell, and A. Haji-Sheikh. Comparison of some inverse heat conduction methods using experimental data. *International Journal of Heat and Mass Transfer*, 39:3649–3657, 1996.

James V. Beck. Nonlinear estimation applied to the nonlinear inverse heat conduction problem. *International Journal of Heat and Mass Transfer*, 13:703–716, 1970.

Julius Berner, Miguel Liu-Schiaffini, Jean Kossaifi, Valentin Duruisseaux, Boris Bonev, Kamyar Azizzadenesheli, and Anima Anandkumar. Principled approaches for extending neural architectures to function spaces for operator learning. *Nature Machine Intelligence*, 8:1173–1181, 2026.

Peter Burgholzer, Michael Thor, Jurgen Gruber, and Gunther Mayr. Three-dimensional thermographic imaging using a virtual wave concept. *Journal of Applied Physics*, 121:105102, 2017.

Shuhao Cao. Choose a transformer: Fourier or galerkin. *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*, 2021.

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*, 4:1155–1164, 2022.

Zhongkai Hao, Zhengyi Wang, Hang Su, Chengyang Ying, Yinpeng Dong, Songming Liu, Ze Cheng, Jian Song, and Jun Zhu. GNOT: A general neural operator transformer for operator learning. In *Proceedings of the 40th International Conference on Machine Learning*, Proceedings of Machine Learning Research, 2023.

Maximilian Herde, Bogdan Raonić, Tobias Rohner, Roger Käppeli, Roberto Molinaro, Emmanuel de Bézenac, and Siddhartha Mishra. Poseidon: Efficient foundation models for pdes. *Advances in Neural Information Processing Systems*, 37:72525–72624, 2024.

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.

Hyunsu Kim, Jonggeon Park, Joan Bruna, Hongseok Yang, and Juho Lee. Axial neural networks for dimension-free foundation models. *Advances in Neural Information Processing Systems*, 38: 112716–112749, 2026.

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*, 24:1–97, 2023a.

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, 2023b.

Thorsten Kurth, Shashank Subramanian, Peter Harrington, Jaideep Pathak, Morteza Mardani, David Hall, Andrea Miele, Karthik Kashinath, and Anima Anandkumar. Fourcastnet: Accelerating global high-resolution weather forecasting using adaptive fourier neural operators. *Proceedings of the Platform for Advanced Scientific Computing Conference*, pp. 1–11, 2023.

Zijie Li, Dule Shu, and Amir Barati Farimani. Scalable transformer for PDE surrogate modeling. In *Advances in Neural Information Processing Systems*, volume 36, 2023a.

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

Zongyi Li, Daniel Zhengyu Huang, Burigede Liu, and Anima Anandkumar. Fourier neural operator with learned deformations for pdes on general geometries. *Journal of Machine Learning Research*, 24:1–26, 2023b.

Lu Lu, Pengzhan Jin, and George Em Karniadakis. Deeponet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators. *arXiv preprint arXiv:1910.03193*, 2019.

Michael McCabe, Bruno Régaldo-Saint Blancard, Liam Parker, Ruben Ohana, Miles Cranmer, Alberto Bietti, Michael Eickenberg, Siavash Golkar, Geraud Krawezik, Francois Lanusse, Mariel Pettee, Tiberiu Tesileanu, Kyunghyun Cho, and Shirley Ho. Multiple physics pretraining for spatiotemporal surrogate models. *Advances in Neural Information Processing Systems*, 37:119301–119335, 2024.

Md Ashiqur Rahman, Zachary E. Ross, and Kamyar Azizzadenesheli. U-no: U-shaped neural operators. *arXiv*, pp. 2204.11127, 2022. doi: https://doi.org/10.1109/LGRS.2018.2802944.

Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378(C), 2018a.

Maziar Raissi, Alireza Yazdani, and George Em Karniadakis. Hidden fluid mechanics: A navier-stokes informed deep learning framework for assimilating flow visualization data. *arXiv*, 2018b. doi: 10.48550/arXiv.1808.04327.

Shashank Subramanian, Peter Harrington, Kurt Keutzer, Wahid Bhimji, Dmitriy Morozov, Michael W. Mahoney, and Amir Gholami. Towards foundation models for scientific machine learning: Characterizing scaling and transfer behavior. *Advances in Neural Information Processing Systems*, 36:71242–71262, 2023.

Jingmin Sun, Yuxuan Liu, Zecheng Zhang, and Hayden Schaeffer. Towards a foundation model for partial differential equations: Multioperator learning and extrapolation. *Physical Review E*, 111: 035304, 2025.

Makoto Takamoto, Timothy Praditia, Raphael Leiteritz, Dan MacKinlay, Francesco Alesiani, Dirk Pflüger, and Mathias Niepert. PDEBench: An extensive benchmark for scientific machine learning. In *Advances in Neural Information Processing Systems: Datasets and Benchmarks Track*, 2022.

Tapas Tripura and Souvik Chakraborty. Wavelet neural operator for solving parametric partial differential equations in computational mechanics problems. *Computer Methods in Applied Mechanics and Engineering*, 404:115783, 2023. doi: https://doi.org/10.1016/j.cma.2022.115783.

Jingbo Wang and Nicholas Zabaras. A bayesian inference approach to the inverse heat conduction problem. *International Journal of Heat and Mass Transfer*, 47:17–18, 2004.

Tian Wang and Chuang Wang. Latent neural operator for solving forward and inverse pde problems. *Advances in Neural Information Processing Systems 37 (NeurIPS 2024)*, 2024. doi: https://doi.org/10.52202/079017-1042.

Shizheng Wen, Arsh Kumbhat, Levi E. Lingsch, Sepehr Mousavi, Yizhou Zhao, Praveen Chandrashekar, and Siddhartha Mishra. Geometry aware operator transformer as an efficient and accurate neural surrogate for pdes on arbitrary domains. *Advances in Neural Information Processing Systems*, 38:155423–155501, 2026.

Huaiqian You, Quinn Zhang, Colton J. Ross, Chung-Hao Lee, and Yue Yu. Learning deep implicit fourier neural operators (ifnos) with applications to heterogeneous material modeling. *Computer Methods in Applied Mechanics and Engineering*, 398:115296, 2022.

Zhengxin Zhang, Qingjie Liu, and Yunhong Wang. Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, 15(5):749–753, 2018. doi: https://doi.org/10.1109/LGRS.2018.2802944.

Pengfei Zhu and Xavier Maldague. Thz-pinns: Time-domain forward modeling of terahertz spectroscopy with physics-informed neural networks. *IEEE Transactions on Terahertz Science and Technology*, 16(6):653–661, 2026.

Pengfei Zhu, Julien Lecompagnon, Philipp Daniel Hirsch, and Mathias Ziegler. Generalized virtual-wave theory for photothermal coherence tomography under arbitrary excitation toward non-contact industrial inspection of composite materials. *arXiv*, 2026a. doi: https://doi.org/10.48550/arXiv.2605.03747.

Pengfei Zhu, Julien Lecompagnon, Philipp Daniel Hirsch, and Mathias Ziegler. A unified framework for virtual wave transform: From generalized formulation to excitation-specific projection. *arXiv*, 2026b. doi: https://doi.org/10.48550/arXiv.2606.08747.

Pengfei Zhu, Julien Lecompagnon, Mathias Ziegler, Clemente Ibarra-Castanedo, and Xavier Maldague. Generalized virtual wave reconstruction for vibrothermography: Overcoming the wavefront-free behavior and quantification challenges in the diffusion-wave field. *arXiv*, 2026c. doi: https://doi.org/10.48550/arXiv.2603.23765.

Pengfei Zhu, Hai Zhang, Stefano Sfarra, Fabrizio Sarasini, Rubén Usamentiaga, Gunther Steenackers, Clemente Ibarra-Castanedo, and Xavier Maldague. Thermal diffusivity characterization of impacted composites using evaporative cryocooling excitation and inverse physics-informed neural networks. *IEEE Transactions on Instrumentation and Measurement*, 75:6003011, 2026d.

A Factorization Proof and Diagnostic Scope

Proof of Proposition 1. Necessity follows because $\mathcal{A}(z_1) = \mathcal{A}(z_2)$ implies $\mathcal{G}_{A\to B}(\mathcal{A}(z_1)) = \mathcal{G}_{A\to B}(\mathcal{A}(z_2))$. For sufficiency, choose any z with $\mathcal{A}(z) = u_A$ and define $\mathcal{G}_{A\to B}(u_A) = \mathcal{B}(z)$. Equation equation 3 makes this definition independent of the representative. Because a continuous surjection from compact $\mathcal{Z}$ to the Hausdorff space $\mathcal{A}(\mathcal{Z})$ is a quotient map, and $\mathcal{B}$ is continuous and constant on each fiber of $\mathcal{A}$, the induced map is continuous. Compactness is used only to establish continuity, not set-theoretic existence. In our experiments, the latent state contains finitely many geometric, material, and excitation parameters restricted to closed bounded ranges, so $\mathcal{Z}$ is modeled as a compact subset of $\mathbb{R}^{d_z}$.

A.1 Physical and Spectral Diagnostics

The present networks predict surface observations and do not receive complete coefficient, source, or boundary fields. Their outputs are therefore not certified PDE solutions. If those quantities are available, target-specific residuals can be evaluated as

$$\mathcal{R}_D[\widehat{T}] = \partial_t \widehat{T} - \nabla\cdot(\alpha\nabla\widehat{T}) - q_D, \qquad \mathcal{R}_W[\widehat{p}] = \partial_{tt}\widehat{p} - \nabla\cdot(c^2\nabla\widehat{p}) - q_W. \tag{19}$$

Residual, boundary, and energy-balance penalties could then be added without guessing hidden physics.

The reported global metrics do not rule out oversmoothing. A complementary diagnostic is

$$E_{\mathrm{spec}} = \frac{\|\mathcal{F}_{x,t}(\widehat{\mathbf{U}}) - \mathcal{F}_{x,t}(\mathbf{U})\|_2}{\|\mathcal{F}_{x,t}(\mathbf{U})\|_2 + \epsilon}, \tag{20}$$

together with band-limited amplitude error and phase error evaluated only above a declared target-magnitude threshold. These diagnostics are future evaluation criteria and are not claimed as reported results.

B Additional Derivation of the Cross-Physics Scaling Principle

This section provides the derivation and physical interpretation of the dimensionless scaling principle introduced in Sec. 3.2. The purpose of the scaling is to construct diffusion and wave observations with comparable characteristic spatial extents over a prescribed observation interval. It does not establish an analytical equivalence between their governing equations.

B.1 Nondimensionalization of Diffusion

Consider the homogeneous diffusion equation

$$\frac{\partial T}{\partial t} = \alpha \nabla^2 T, \tag{21}$$

where α is the thermal diffusivity. We introduce characteristic spatial and temporal scales L and t_0, together with a reference field amplitude T_0,

$$\mathbf{x} = L\boldsymbol{\xi}, \qquad t = t_0 s, \qquad T = T_0 T^*. \tag{22}$$

The temporal and spatial derivatives become

$$\frac{\partial T}{\partial t} = \frac{T_0}{t_0}\frac{\partial T^*}{\partial s}, \qquad \nabla^2 T = \frac{T_0}{L^2}\nabla_\xi^2 T^*. \tag{23}$$

Substituting these expressions into Eq. equation 21 yields

$$\frac{\partial T^*}{\partial s} = \underbrace{\frac{\alpha t_0}{L^2}}_{\mathrm{Fo}} \nabla_\xi^2 T^*, \tag{24}$$

where $\mathrm{Fo} = \alpha t_0 / L^2$ is the Fourier number. The characteristic diffusion time across a distance L is $\tau_D = L^2/\alpha$. Consequently, $\mathrm{Fo} = t_0/\tau_D$. The Fourier number therefore describes the observation duration relative to the characteristic diffusion time.

B.2 Characteristic Diffusion Length

For an impulsive initial condition in an unbounded homogeneous medium, the d-dimensional diffusion Green's function is

$$G_D(\mathbf{x}, t) = \frac{1}{(4\pi\alpha t)^{d/2}} \exp\left(-\frac{|\mathbf{x}|^2}{4\alpha t}\right), \qquad t > 0. \tag{25}$$

The exponential factor identifies the characteristic diffusion length, $\ell_D(t) \sim \sqrt{\alpha t}$, up to a numerical factor determined by the selected definition of penetration depth. At $t = t_0$, its normalized value is $\ell_D(t_0)/L \sim \sqrt{\mathrm{Fo}}$. Thus, $\sqrt{\mathrm{Fo}}$ has a geometric interpretation as the characteristic diffusion length normalized by the system size. This interpretation is particularly useful for constructing cross-physics training data because it relates a dimensionless PDE coefficient to the spatial extent explored during the observation interval.

B.3 Nondimensionalization of Wave Propagation

Consider the homogeneous wave equation

$$\frac{\partial^2 p}{\partial t^2} = c^2 \nabla^2 p, \tag{26}$$

where c denotes the wave speed. Introduce the same spatial and temporal scales L and t_0,

$$\mathbf{x} = L\boldsymbol{\xi}, \qquad t = t_0 s, \qquad p = p_0 p^*. \tag{27}$$

Substitution into Eq. equation 26 gives

$$\frac{\partial^2 p^*}{\partial s^2} = \underbrace{\left(\frac{ct_0}{L}\right)^2}_{\Lambda^2} \nabla_\xi^2 p^*, \tag{28}$$

where $\Lambda = ct_0/L$. The characteristic wave transit time is $\tau_W = L/c$. Therefore, $\Lambda = t_0/\tau_W$. Unlike diffusion, wave propagation has a characteristic travel distance $\ell_W(t) = ct$. At the observation time t_0, $\ell_W(t_0)/L = \Lambda$.

B.4 Matching Characteristic Spatial Scales

We construct diffusion–wave pairs by matching their characteristic spatial extents over the same observation time, $\ell_D(t_0) \sim \ell_W(t_0)$. Using $\ell_D(t_0)/L \sim \sqrt{\mathrm{Fo}}$ and $\ell_W(t_0)/L = \Lambda$, this condition becomes $\sqrt{\mathrm{Fo}} \sim \Lambda$. Squaring both sides yields $\mathrm{Fo} \sim \Lambda^2$ We select the representative equality $\mathrm{Fo} = \Lambda^2$ as the default data-generation constraint. Substituting the definitions of Fo and Λ gives

$$\frac{\alpha t_0}{L^2} = \frac{c^2 t_0^2}{L^2}. \tag{29}$$

Hence, $c = \sqrt{\alpha/t_0}$ This expression defines a cross-domain scaling parameter. The selected value of c depends explicitly on the observation time t_0. It must therefore not be interpreted as a finite thermal propagation speed derived from Fourier heat conduction.

B.5 Generalized Scale Alignment

Exact equality of the characteristic lengths is not necessary. A generalized alignment factor $\chi > 0$ can be introduced through $\ell_W(t_0) = \chi \ell_D(t_0)$. This gives $\Lambda = \chi\sqrt{\mathrm{Fo}}$, or equivalently,

$$\Lambda^2 = \chi^2 \mathrm{Fo} \tag{30}$$

The corresponding wave speed is

$$c = \chi\sqrt{\frac{\alpha}{t_0}}. \tag{31}$$

The default configuration uses $\chi = 1$. The generalized formulation provides a parameterization for future studies of scale mismatch. The present experiments do not constitute a systematic ablation over χ.

B.6 Why Scaling Does Not Imply Dynamical Equivalence

The distinction between diffusion and wave propagation becomes particularly clear in spectral space. For homogeneous diffusion, spatial Fourier transformation gives

$$\frac{\partial \widehat{T}}{\partial t} = -\alpha |\mathbf{k}|^2 \widehat{T}. \tag{32}$$

The solution is

$$\widehat{T}(\mathbf{k}, t) = e^{-\alpha |\mathbf{k}|^2 t} \widehat{T}(\mathbf{k}, 0). \tag{33}$$

For the homogeneous wave equation,

$$\frac{\partial^2 \widehat{p}}{\partial t^2} + c^2 |\mathbf{k}|^2 \widehat{p} = 0. \tag{34}$$

Its general solution is

$$\widehat{p}(\mathbf{k}, t) = A_{\mathbf{k}} e^{ic|\mathbf{k}|t} + B_{\mathbf{k}} e^{-ic|\mathbf{k}|t}. \tag{35}$$

The diffusion multiplier produces exponential attenuation, whereas the ideal wave multipliers describe oscillatory phase evolution. Matching $\mathrm{Fo} = \Lambda^2$ does not make these multipliers equal. The proposed scaling therefore aligns characteristic observation scales, rather than identifying the two governing operators.

B.7 Extension to Heterogeneous Media

For heterogeneous materials, the coefficients depend on position,

$$\frac{\partial T}{\partial t} = \nabla \cdot [\alpha(\mathbf{x}) \nabla T], \quad \frac{\partial^2 p}{\partial t^2} = \nabla \cdot \left[c^2(\mathbf{x}) \nabla p\right]. \tag{36}$$

In general, these systems do not possess a single global diffusion length or wave speed. We therefore introduce reference background coefficients α_b and c_b,

$$\mathrm{Fo}_b = \frac{\alpha_b t_0}{L^2}, \qquad \Lambda_b = \frac{c_b t_0}{L}. \tag{37}$$

The reference scaling condition is $\mathrm{Fo}_b = \Lambda_b^2$. The local material coefficients are represented as

$$\alpha(\mathbf{x}) = \alpha_b a(\mathbf{x}), \qquad c(\mathbf{x}) = c_b q(\mathbf{x}). \tag{38}$$

Thus, the global scaling controls the background evolution regime, while the heterogeneous coefficient fields determine the distinct responses of each governing operator to the shared structure. The scaling condition is not imposed pointwise throughout the heterogeneous material.

C Numerical Simulation and Dataset Details

This section provides additional information about the numerical construction of the paired diffusion–wave dataset.

C.1 Computational Domain

The full physical fields are generated over a two-dimensional rectangular domain,

$$\Omega = [0, L_x] \times [0, L_y], \quad L_x = 50.0\ \mathrm{mm}, \qquad L_y = 5.0\ \mathrm{mm}. \tag{39}$$

The numerical grid contains

$$N_x \times N_y = 5001 \times 501 \quad \Delta x = \frac{L_x}{N_x - 1}, \qquad \Delta y = \frac{L_y}{N_y - 1}. \tag{40}$$

The two PDEs are solved on the full spatial domain, but only boundary observations are provided to the learning models.

C.2 Latent Structural Model

The latent variables are divided into geometry, excitation, and material parameters,

$$z = (\boldsymbol{\theta}_g, \boldsymbol{\theta}_s, \boldsymbol{\theta}_p), \quad \boldsymbol{\theta}_g = \{x_{c,j}, y_{c,j}, a_j, b_j, \psi_j, g_j\}_{j=1}^{N_d}, \quad \boldsymbol{\theta}_p = (\alpha_b, c_b, \delta_\alpha, \delta_c). \tag{41}$$

Here g_j specifies the geometric family, such as an ellipse or rectangle, while δ_α and δ_c independently specify the modality-dependent material contrasts. For defect j, let (x'_j, y'_j) denote coordinates rotated by ψ_j. Elliptical inclusions are represented by

$$r^{(2)}_{d,j}(x,y) = \left(\frac{x'_j}{a_j}\right)^2 + \left(\frac{y'_j}{b_j}\right)^2, \tag{42}$$

whereas rectangular inclusions use

$$r^{(\infty)}_{d,j}(x,y) = \max\left(\frac{|x'_j|}{a_j}, \frac{|y'_j|}{b_j}\right). \tag{43}$$

A square is obtained by setting $a_j = b_j$. In either case, the corresponding smooth mask is

$$m_j(x,y) = \frac{1}{2}\left[1 - \tanh\left(\frac{r_{d,j}(x,y) - 1}{\varepsilon_d}\right)\right]. \tag{44}$$

For multiple defects, the union mask is

$$m(x,y) = 1 - \prod_{j=1}^{N_d} \left[1 - m_j(x,y)\right]. \tag{45}$$

Both fields therefore respond to the same geometric configuration, but their material contrasts enter different physical evolution equations.

C.3 Diffusion Solver

The diffusion field satisfies

$$\frac{\partial T}{\partial t} = \nabla \cdot [\alpha(x,y)\nabla T] + s_D(x,y,t). \tag{46}$$

Surface excitation is imposed through the boundary condition

$$-\mathbf{n} \cdot [\alpha(x,y)\nabla T] = q_D(x,t) \qquad \text{on } \Gamma_s, \tag{47}$$

where Γ_s denotes the accessible boundary. The remaining boundaries (homogeneous Neumann conditions) and initial condition are

$$\mathbf{n} \cdot \nabla T = 0 \qquad \text{on } \partial\Omega \setminus \Gamma_s. \quad T(x,y,0) = 0. \tag{48}$$

The equation is discretized using second-order spatial finite differences. For explicit time marching, the diffusion time step satisfies

$$\Delta t_D \le \frac{1}{2\alpha_{\max}\left(\Delta x^{-2} + \Delta y^{-2}\right)}, \quad \alpha_{\max} = \max_{(x,y)\in\Omega} \alpha(x,y). \tag{49}$$

C.4 Wave Solver

The independently generated wave field satisfies

$$\frac{\partial^2 p}{\partial t^2} = \nabla \cdot \left[c^2(x,y)\nabla p\right] + s_W(x,y,t). \tag{50}$$

The initial conditions are

$$p(x,y,0) = 0, \qquad \frac{\partial p}{\partial t}(x,y,0) = 0. \tag{51}$$

For the reported simulations, the interior source is set to $s_W(x,y,t) = 0$, and excitation is imposed through the Neumann boundary condition

$$-\mathbf{n} \cdot \left[c^2(x,y)\nabla p\right] = q_W(x,t) \qquad \text{on } \Gamma_s. \tag{52}$$

The remaining boundaries are reflecting,

$$\mathbf{n} \cdot \nabla p = 0 \qquad \text{on } \partial\Omega \setminus \Gamma_s. \tag{53}$$

The wave equation is discretized using second-order finite differences in space and time. The numerical time step is selected according to the two-dimensional CFL condition,

$$c_{\max} \Delta t_W \sqrt{\Delta x^{-2} + \Delta y^{-2}} \leq C_{\mathrm{CFL}}, \quad C_{\mathrm{CFL}} \leq 1. \tag{54}$$

The diffusion and wave solvers therefore use independently selected internal time steps appropriate to their respective numerical schemes.

C.5 Boundary Observations and Data Pairing

Let $\Gamma_s = \{(x, y) \in \Omega : y = 0\}$ denote the observed surface. We extract boundary observations using $\mathcal{O}_\Gamma f = f|_{\Gamma_s}$. For latent configuration z_i, the paired observations are

$$\begin{aligned} \mathbf{U}_D^{(i)} &= \mathcal{O}_\Gamma \mathcal{S}_D(z_i), \\ \mathbf{U}_W^{(i)} &= \mathcal{O}_\Gamma \mathcal{S}_W(z_i). \end{aligned} \tag{55}$$

The resulting fields are resampled onto a common observation grid,

$$\mathbf{U}_D^{(i)}, \mathbf{U}_W^{(i)} \in \mathbb{R}^{501 \times 200}. \tag{56}$$

This resampling is performed after the independent PDE simulations. The input and target fields therefore have the same observational dimensions without requiring their numerical solvers to use identical internal grids or time steps.

C.6 Dataset Partition and Normalization

The complete dataset contains $N = 1000$ paired observations. The reported partition is

$$N_{\mathrm{train}} = 800, \qquad N_{\mathrm{val}} = 100, \qquad N_{\mathrm{test}} = 100. \tag{57}$$

Paired observations originating from the same latent configuration are assigned to the same subset. The two modalities are standardized independently,

$$\widetilde{\mathbf{U}}_j = \frac{\mathbf{U}_j - \mu_j}{\sigma_j}, \qquad j \in \{D, W\}. \tag{58}$$

All normalization statistics are computed from the training subset and are subsequently applied unchanged to the validation and test subsets.

D Neural Network Implementation Details

We provide additional details for the seven architectures evaluated in Sec. 4. All networks operate on single-channel source fields sampled on the 501×200 observation grid and produce target fields at the same resolution. The prediction direction determines whether the input and target correspond to diffusion or wave propagation.

D.1 ResUNet

ResUNet Zhang et al. (2018) is implemented as a residual encoder–decoder with four resolution levels (see Figure 6). The encoder feature widths are

$$32 \to 64 \to 128 \to 256, \tag{59}$$

followed by a 512-channel bottleneck. Each encoder stage contains a residual convolutional block and 2×2 downsampling. The decoder reverses the feature hierarchy,

$$512 \to 256 \to 128 \to 64 \to 32. \tag{60}$$

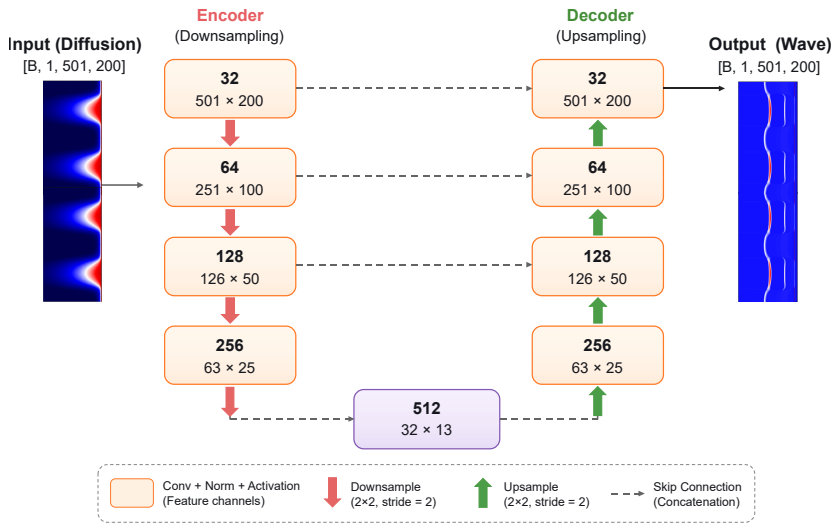

Figure 6: ResUNet architecture for diffusion-to-wave cross-physics mapping. The network maps a spatiotemporal diffusion field $T(x,t) \in \mathbb{R}^{501\times200}$ to the corresponding wave field $p(x,t) \in \mathbb{R}^{501\times200}$. The encoder progressively extracts multiscale representations through convolutional residual blocks and 2×2 downsampling, increasing the feature dimension from 32 to 512 channels while reducing the spatiotemporal resolution. The decoder symmetrically restores the original resolution through successive upsampling stages. Skip connections concatenate encoder features with the corresponding decoder features at matched scales, preserving fine-scale spatiotemporal information that may otherwise be lost in the latent representation. The bottleneck provides a compact high-dimensional representation connecting the diffusive and propagative physical domains. Numbers inside each block indicate the number of feature channels and the corresponding $(T \times X)$ feature-map resolution.

Encoder–decoder skip connections transfer features between corresponding resolution levels. The decoder features are interpolated to the exact resolution of the relevant encoder features before concatenation. This explicit resizing is important because the temporal dimension 501 is not divisible by 2^4. A final 1×1 convolution produces the single-channel output.

D.2 U-Shaped Neural Operator

U-NO Rahman et al. (2022) combines spectral operator blocks with a multiresolution U-shaped architecture (see Figure 7). The input field is augmented with normalized temporal and spatial coordinates. A pointwise lifting layer maps the three input channels to a latent width of 32. The encoder feature widths are

$$32 \to 64 \to 128 \to 256. \tag{61}$$

The retained Fourier mode pairs are

$$(M_t, M_x) = \begin{cases} (64, 40), \\ (32, 24), \\ (16, 12), \\ (8, 6) \end{cases} \tag{62}$$

across the successive encoder and bottleneck levels. Each operator block combines a truncated Fourier transformation with a learned pointwise 1×1 convolution, group normalization, and GELU activation. The encoder uses average pooling with `ceil_mode=True` to accommodate the odd temporal dimension. The decoder progressively restores the original resolution using interpolation and skip connections. A local residual refinement module and a pointwise projection generate the single-channel target field.

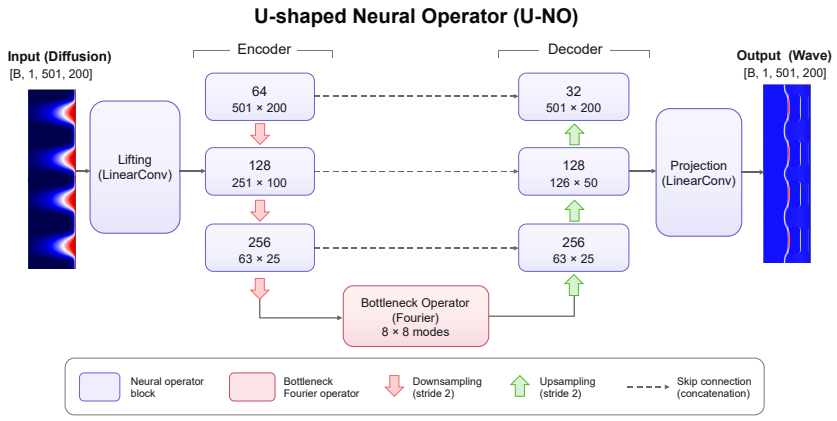

Figure 7: U-Shaped Neural Operator (U-NO) for diffusion-to-wave cross-physics mapping. The model learns the operator mapping from an input diffusion field $T(x,t) \in \mathbb{R}^{501\times200}$ to the corresponding wave field $p(x,t) \in \mathbb{R}^{501\times200}$. A lifting layer first embeds the scalar input field into a higher-dimensional feature representation, which is progressively compressed by the U-shaped encoder to construct multiscale latent features. At the coarsest representation, a Fourier neural operator with 8×8 retained modes captures nonlocal interactions in the joint space–time domain. The decoder subsequently restores the original spatiotemporal resolution, while skip connections transfer multiscale encoder features to the corresponding decoder stages. Finally, a linear projection maps the decoded representation back to the physical wave field. Numbers inside the blocks denote feature channels and $(T \times X)$ feature-map resolutions, respectively.

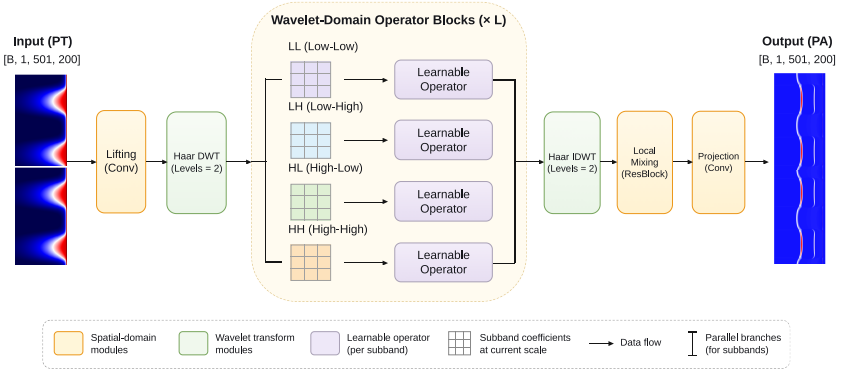

Figure 8: Wavelet Neural Operator (WNO) for cross-physics operator learning. Given an input diffusion field $T(x,t) \in \mathbb{R}^{501\times200}$, WNO predicts the corresponding wave field $p(x,t) \in \mathbb{R}^{501\times200}$ through a multiresolution wavelet representation. The input is first lifted to a higher-dimensional feature space and decomposed using a two-level Haar discrete wavelet transform (DWT). The resulting low–low (LL), low–high (LH), high–low (HL), and high–high (HH) subbands are processed by parallel learnable operator branches, allowing the model to capture cross-physics interactions at distinct spatiotemporal scales and frequency bands. After L wavelet-domain operator blocks, the transformed features are reconstructed by the inverse discrete wavelet transform (IDWT). A residual local-mixing block further refines short-range features before a final projection maps the representation to the physical wave field. This multiresolution construction provides an explicit mechanism for jointly representing smooth diffusive structures and rapidly varying propagative features within the learned diffusion-to-wave operator.

D.3 Wavelet Neural Operator

WNO Tripura & Chakraborty (2023) lifts the input and coordinate channels to a latent width of 48 (see Figure 8). Its operator blocks use a two-level Haar discrete wavelet transform. At each

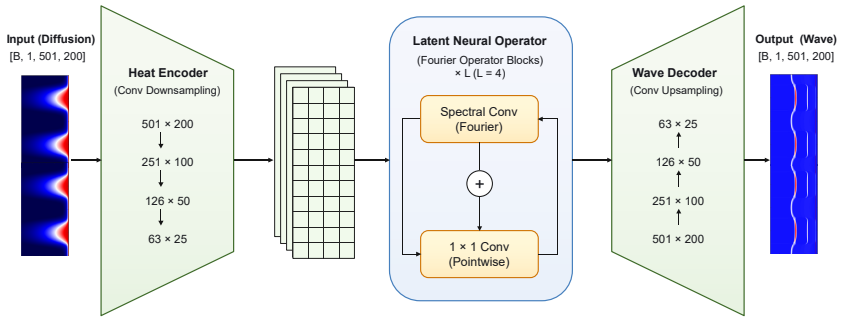

Figure 9: Latent Neural Operator (LNO) for cross-physics operator learning. Given an input diffusion field $T(x,t) \in \mathbb{R}^{501\times200}$, LNO predicts the corresponding wave field $p(x,t) \in \mathbb{R}^{501\times200}$ by learning the cross-physics transformation in a compressed latent space. A convolutional heat encoder progressively downsamples the input field from 501×200 to 63×25, producing a compact multichannel latent representation. The latent field is then processed by $L = 4$ Fourier operator blocks, each combining a spectral convolution for nonlocal interactions with a pointwise 1×1 convolution for local feature mixing. The resulting latent representation is mapped back to the physical domain by a convolutional wave decoder that progressively restores the original spatiotemporal resolution. By performing operator learning on the compressed representation, LNO provides an efficient mechanism for capturing global diffusion-to-wave dependencies without applying spectral operations directly at the full field resolution.

decomposition level, the representation is separated into four subbands,

$$\mathcal{W}(\mathbf{z}) = (\mathbf{z}_{LL}, \mathbf{z}_{LH}, \mathbf{z}_{HL}, \mathbf{z}_{HH})\,. \tag{63}$$

Each subband is processed by an independently learned 1×1 channel-mixing operator. The transformed coefficients are reconstructed using the inverse discrete wavelet transform. Each WNO block additionally contains a pointwise convolution and local residual refinement. Four WNO blocks are stacked sequentially at a fixed latent width of 48. Replication padding is applied before wavelet decomposition when the feature dimensions are odd. The reconstructed field is cropped to its original resolution after inverse transformation.

D.4 Latent Neural Operator

As in U-NO and FNO, the single-channel source field is concatenated with normalized temporal and spatial coordinate channels, producing the three-channel encoder input. LNO Wang & Wang (2024) first compresses the input field into a latent representation before applying spectral operator learning (see Figure 9). The encoder channel widths are

$$3 \to 32 \to 64 \to 96 \to 128. \tag{64}$$

The observation grid is progressively reduced,

$$501 \times 200 \to 251 \times 100 \to 126 \times 50 \to 63 \times 25. \tag{65}$$

The resulting latent representation has dimensions

$$\mathbf{z} \in \mathbb{R}^{128\times63\times25}. \tag{66}$$

Four Fourier operator blocks are applied within this compressed representation. The nominal retained mode pair is

$$(M_t, M_x) = (24, 12). \tag{67}$$

An additional residual pointwise transformation maps the processed latent features toward the target modality. The decoder restores the original resolution through successive interpolation and residual convolutional blocks. Unlike ResUNet and U-NO, the LNO decoder does not use encoder–decoder skip connections.

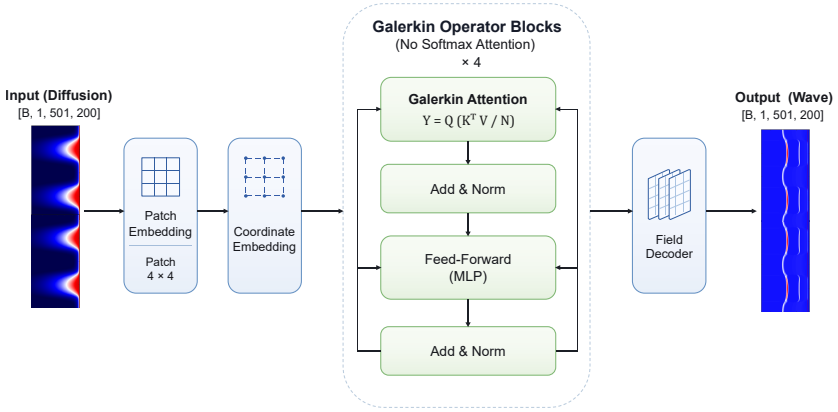

Figure 10: Galerkin Neural Operator (GNO) for cross-physics operator learning. Given an input diffusion field $T(x,t) \in \mathbb{R}^{501\times200}$, GNO predicts the corresponding wave field $p(x,t) \in \mathbb{R}^{501\times200}$ through a sequence of Galerkin operator blocks. The input field is first partitioned into 4×4 patches and mapped to latent tokens, which are augmented with coordinate embeddings to retain spatiotemporal positional information. Four Galerkin operator blocks then model global interactions among the latent field representations. Unlike conventional softmax attention, the Galerkin attention evaluates $Y = Q(K^\top V/N)$ directly, providing an operator-oriented global mixing mechanism without softmax normalization. Each block combines Galerkin attention with residual normalization and a feed-forward network. Finally, a field decoder reconstructs the full-resolution wave field from the processed latent representation. This architecture provides a global, coordinate-aware alternative to convolutional and spectral operator representations for learning the diffusion-to-wave transformation.

D.5 Galerkin Neural Operator

GNO Cao (2021) employs a patch-based Galerkin attention architecture (see Figure 10). The input field is replication-padded along the temporal dimension,

$$501 \times 200 \to 504 \times 200. \tag{68}$$

Using non-overlapping 4×4 patches produces

$$N = 126 \times 50 = 6300 \tag{69}$$

tokens. The embedding dimension is $d = 96$. Each token is augmented with normalized coordinate information. For query, key, and value representations $\mathbf{Q}$, $\mathbf{K}$, and $\mathbf{V}$, the Galerkin interaction is

$$\mathcal{A}_G(\mathbf{Q},\mathbf{K},\mathbf{V}) = \mathbf{Q}\left(\frac{\mathbf{K}^\top\mathbf{V}}{N}\right). \tag{70}$$

The implementation uses four attention heads, each with dimension 24. Four Galerkin operator blocks are stacked with residual connections and feed-forward layers. The output tokens are reshaped into a latent patch field and refined using local convolutions. A transposed convolution restores the full-resolution field, after which the additional temporal padding is removed.

D.6 DeepONet

DeepONet Lu et al. (2019) represents the cross-physics operator through separate branch and trunk networks. The branch network encodes the complete source observation into an input-dependent coefficient vector. It uses convolutional feature widths

$$1 \to 32 \to 64 \to 128 \to 256 \to 256. \tag{71}$$

Adaptive average pooling compresses the final convolutional representation to a 4×4 feature map. The pooled $256 \times 4 \times 4$ tensor is flattened into a 4096-dimensional vector and projected by a learned

affine map from $\mathbb{R}^{4096}$ to $\mathbb{R}^{128}$. The branch output dimension is 128. The trunk network maps normalized coordinates (t, x) to 128 coordinate-dependent basis functions. The predicted field is obtained from the branch–trunk contraction,

$$\widehat{u}_B(t,x) = \frac{1}{\sqrt{p}} \sum_{k=1}^{p} b_k(\mathbf{U}_A)\phi_k(t,x) + b_0, \tag{72}$$

where $p = 128$. The output is evaluated on the full observation grid and reshaped to 501×200.

D.7 Fourier Neural Operator

FNO Li et al. (2020) operates directly on the full-resolution spatiotemporal field. The input is augmented with normalized temporal and spatial coordinates, then lifted to a latent width of 64. The network contains four Fourier operator blocks. Each block combines a spectral convolution and a pointwise 1×1 transformation,

$$\mathbf{z}_{\ell+1} = \phi\left[\mathcal{K}_\ell(\mathbf{z}_\ell) + \mathcal{W}_\ell(\mathbf{z}_\ell)\right], \tag{73}$$

where ϕ denotes GELU. The retained Fourier mode pair is

$$(M_t, M_x) = (96, 48). \tag{74}$$

After four operator blocks, two pointwise projection layers map the latent representation according to

$$64 \to 128 \to 1. \tag{75}$$

The model preserves the observation resolution throughout its spectral operator layers.

D.8 Training Objectives and Optimization

All seven architectures use an MSE–relative-error reconstruction objective,

$$\mathcal{L}_{\text{rec}} = \mathcal{L}_{\text{MSE}} + 0.1\mathcal{L}_{\text{rel}}, \tag{76}$$

where the relative error is stabilized using

$$\epsilon = 10^{-8}. \tag{77}$$

For FNO with a wave target, the reported objective additionally includes temporal and spatial gradient discrepancies,

$$\mathcal{L}_{\text{FNO}} = \mathcal{L}_{\text{rec}} + 0.02\mathcal{L}_t + 0.02\mathcal{L}_x. \tag{78}$$

An additional Fourier-domain loss was considered in the original implementation, but its weight was set to zero for the reported experiments. The reported optimization configuration uses AdamW with learning rate $= 10^{-3}$, weight decay $= 10^{-5}$, cosine learning-rate annealing, and gradient clipping at a maximum norm of 1.0. Model selection uses the minimum validation relative ℓ_2 error. The FNO-specific objective should be considered when interpreting comparisons between architectures: the benchmark does not isolate the effect of network design alone.

Figure 11 compares the training and validation trajectories of the seven reported configurations in both mapping directions.

E Additional Qualitative Results

This section provides additional interpretation of the reconstruction examples involving chirped and spatially nonuniform excitations. The corresponding original visualizations are Figure 12 and 13.

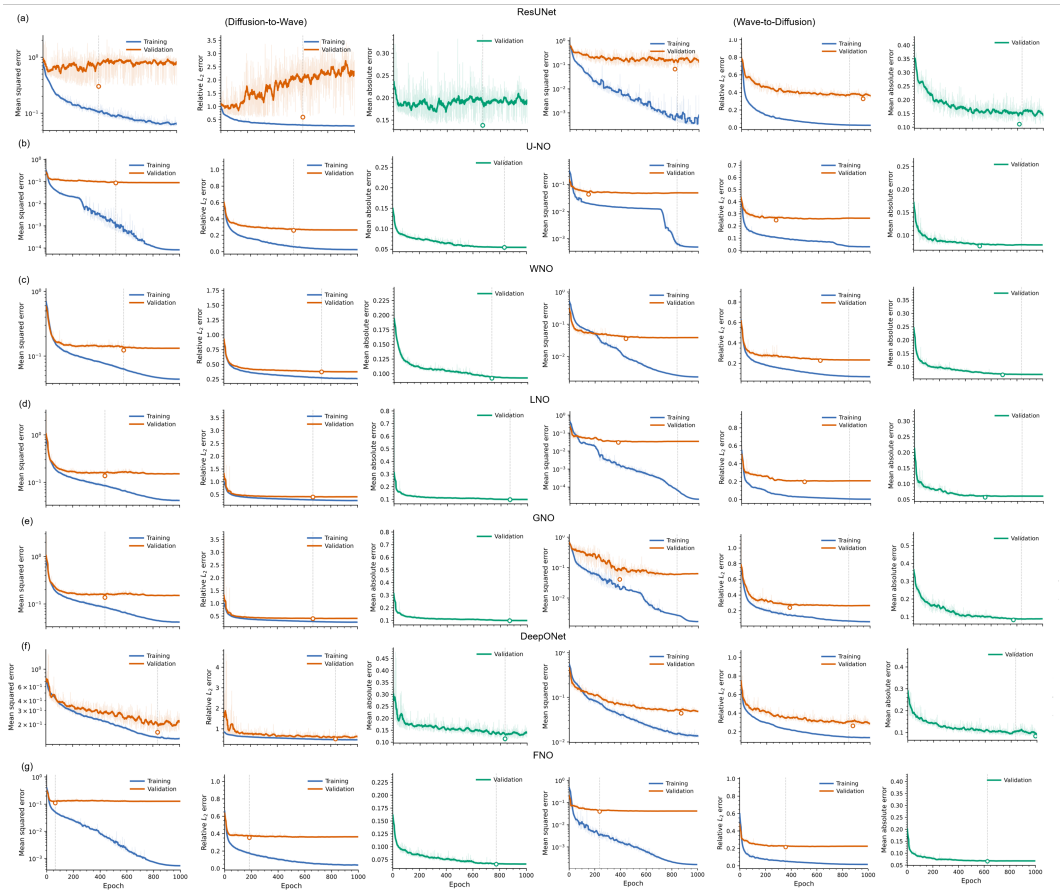

Figure 11: Training and validation trajectories for both mapping directions.

E.1 Defect-Free Chirp Excitation

The defect-free chirp configuration isolates excitation-dependent cross-physics structures without introducing defect scattering (see Figure 12). In the diffusion-to-wave direction, the target contains coherent high-amplitude structures whose locations and temporal organization must be reconstructed from a smoother diffusion observation. The original qualitative comparison shows that U-NO reproduces the dominant wave response with comparatively weak residuals. FNO and LNO also capture the principal structures, but coherent discrepancies remain near some of the high-amplitude features. WNO and GNO recover the large-scale organization with more pronounced structured errors. ResUNet and DeepONet exhibit stronger smoothing or additional low-frequency artifacts. In the reverse direction, the neural operators recover the dominant smooth diffusion structures, although localized amplitude and temporal errors remain. This experiment illustrates that the learned correspondence is not restricted to the recognition of defect-induced scattering patterns.

E.2 Gaussian-Chirp Excitation

The spatially nonuniform Gaussian-chirp configuration introduces localized excitation and an uneven amplitude distribution (see Figure 13). The target wave field contains several localized high-amplitude responses together with weaker distributed components. U-NO reconstructs the dominant response locations and amplitudes with comparatively small residuals. FNO and LNO recover the principal structures but retain more visible errors near localized features. WNO also captures the main response patterns, whereas ResUNet and DeepONet show stronger smoothing and reconstruction artifacts. The reverse mapping produces comparatively smooth target fields. Several operator-

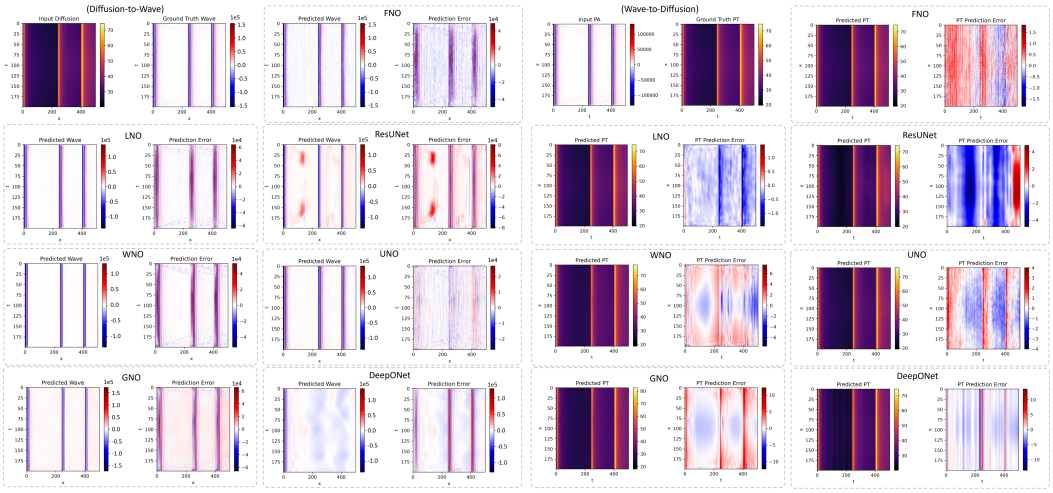

Figure 12: Qualitative comparison of diffusion-to-wave (left) and wave-to-diffusion (right) reconstruction for a defect-free plate under chirp excitation. The diffusion/wave input, ground-truth wave/diffusion field, and predictions obtained using FNO, LNO, ResUNet, WNO, U-NO, GNO, and DeepONet are shown together with the corresponding prediction errors.

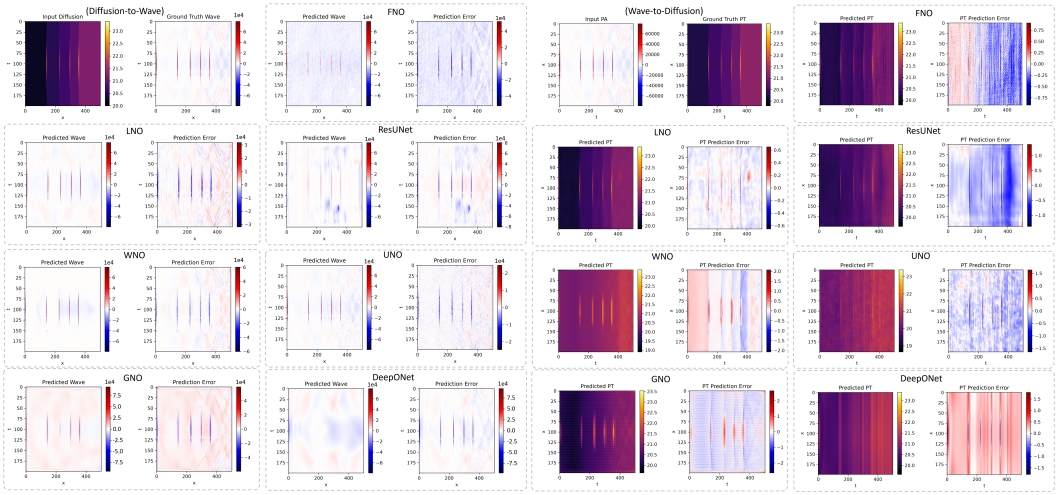

Figure 13: Qualitative comparison of diffusion-to-wave (left) and wave-to-diffusion (right) reconstruction for the case of a centered Gaussian-chirp spatially nonuniform excitation. The diffusion/wave input, ground-truth wave/diffusion field, and predictions obtained using FNO, LNO, ResUNet, WNO, U-NO, GNO, and DeepONet are shown together with their corresponding prediction errors.

based architectures recover their principal spatiotemporal structures, but model-dependent localized residuals remain. The experiment therefore tests the ability of CPM to represent changes in excitation location and amplitude distribution rather than only changes in defect geometry.

F Scope and Limitations

We conclude by clarifying the scope of the theoretical and experimental findings.

F.1 Existence Versus Identifiability

A shared latent configuration does not guarantee that one physical observation uniquely determines another.

As established in Sec. 3.1, a deterministic mapping requires consistency of the target observations on the fibers of the source observation operator.

If two latent configurations produce the same source observation but different target fields, no single-valued deterministic operator can exactly recover both targets.

The synthetic dataset provides paired realizations for supervised learning, but its reported reconstruction errors do not establish global identifiability over arbitrary latent configurations.

F.2 Information Loss and Statistical Prediction

Diffusive measurements attenuate fine-scale spatial information, whereas ideal wave propagation preserves oscillatory phase evolution.

Consequently, diffusion-to-wave prediction may require estimating target components that are weakly represented in the source observation.

A supervised neural network can exploit regularities in the distribution of training configurations, but accurate prediction does not imply exact recovery of information absent from its input.

When the deterministic mapping is ambiguous, the learned output should be interpreted as a conditional statistical estimate.

F.3 Experimental Scope

The reported experiments use paired synthetic fields generated from numerical simulations.

They establish reconstruction performance for the considered configuration distribution and test partition.

They do not establish the same accuracy for arbitrary experimental noise, unseen material classes, different geometries, or substantially different observation conditions.

The scaling principle also remains a data-construction criterion rather than a demonstrated guarantee of identifiability or improved learning performance.

Systematic scale ablations, uncertainty quantification, and physical experimental validation are therefore important directions for subsequent work.

F.4 Baseline Coverage and Objective Fairness

The seven reported rows do not exhaust modern operator-learning architectures. In particular, GNOT and FactFormer are absent, and the standard PDEBench tasks evaluate within-physics surrogates rather than paired cross-PDE translation. The results therefore establish a CPM benchmark instance, not comprehensive state-of-the-art superiority.

In addition, the D2W FNO configuration uses temporal and spatial gradient losses that are absent from the other configurations. Its score reflects both its architecture and its supervision. We retain the row for reproducibility, mark it explicitly in the table, and do not use it to make an architecture-only claim. A future controlled study should cross the factors $\{$architecture$\}$ and $\{$common versus gradient-augmented objective$\}$ for every model.

F.5 Physical and Spectral Certification

Because only surface fields are supplied to the networks, the present predictions are not accompanied by a certificate of target-PDE residual, conservation, or thermodynamic admissibility. Such a certificate requires the coefficient, source, and boundary data that are hidden in the current observation-only task. Likewise, the five reported global metrics do not independently verify high-frequency

amplitude or phase. Frequency-banded amplitude error, phase error, target-PDE residuals, and energy-balance diagnostics should accompany future results when the required physical variables are available. Until then, the paper's conclusions concern conditional field reconstruction under the stated data distribution rather than strict physical equivalence or conservation.