FINITE-PROPAGATION-AWARE MULTIRESOLUTION NEURAL OPERATORS FOR STRUCTURAL DYNAMICS

Anonymous authors
Paper under double-blind review

ABSTRACT

Learning structural-dynamics surrogate models requires reasoning over irregular meshes and sparse sensor measurements while respecting the finite wave-propagation time for loads to arrive at different spatial locations. Standard neural operators can capture long-range spatial interactions, yet unconstrained message passing can produce unphysical early responses that violate wave-causality constraints. We introduce FPMR-NO, a finite-propagation-aware multiresolution neural operator that combines causal load-history encoding, wave travel-time gates, and coarse-to-fine graph aggregation. The model propagates separate latent messages corresponding to primary/compressional waves (P-waves), secondary/shear waves (S-waves), and mixed wave modes across graph edges only once the source-to-edge wave arrival condition is satisfied. This multiresolution pooling shortens computational paths without introducing unphysical zero-time shortcuts. On the DeformingPlate stress task, FPMR-NO achieves the lowest mean normalized relative L2 error among the evaluated methods, with a 2.4% relative improvement over AMG and a 1.7% improvement over Transolver. On real ETH four-story impact tests, it reduces macro-channel normalized root mean squared error (NRMSE) to 0.86, a 14.1% error reduction relative to the strongest baseline. Controlled graph-wave experiments further show a 71.8% reduction in relative L2 error and a 13.3-fold reduction in pre-arrival energy against an ungated counterpart.

1 INTRODUCTION

Structural-dynamics surrogate modeling uses measured loads, geometry, and sensor responses to learn an operator mapping from excitation and structural state to future response. This paradigm builds upon well-established operator-learning and mesh-based simulation frameworks (Lu et al., 2021; Kovachki et al., 2023; Pfaff et al., 2021). During inference, the trained operator can replace repeated numerical simulation or dense sensor processing with a fast response estimate. Consequently, this problem setting demands more than mere approximation accuracy. The operator must handle irregular meshes or sparse sensors, propagate disturbances across long spatial and temporal ranges, and preserve the finite time required for an impact or load to reach each response location.

Neural operators have demonstrated remarkable capacity for learning function-space mappings governed by partial differential equations (PDEs) on regular grids (Li et al., 2021). Graph-based simulators and geometry-aware operators further extend this progress to arbitrary domains, unstructured meshes, and learned geometric representations (Pfaff et al., 2021; Li et al., 2025; Wu et al., 2024; Mousavi et al., 2025). These models show that learned surrogates can capture complex spatial interactions without hand-designing a new discretization rule for every geometry. Despite these advances, existing graph-based neural-operator frameworks suffer a fundamental modeling mismatch that computational connectivity in the learned graph does not equate to physical reachability. Most architectures allow information to move through the learned computation graph as soon as an edge exists. In static or quasi-static tasks, this immediate computational connectivity can be a useful inductive bias, because long-range dependencies are often the target. In structural dynamics, however, the same bias can become an unphysical shortcut if propagation time is not explicitly constrained. Distant nodes can participate in updates before load information would physically arrive. This violates wave-causality and can degrade generalization, especially for impact-driven scenarios dominated by elastic P- and S-wave propagation.

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.

In this work, we argue that finite wave propagation should be embedded as an explicit architectural prior for structural-dynamics neural operators, rather than being entirely left for models to infer from limited training data. Such a physical prior does not replace data-driven learning, and instead, it imposes constraints on when learned latent messages can contribute to node updates. To realize this idea, we propose Finite-Propagation-Aware Multiresolution Neural Operator (FPMR-NO), a graph-based neural operator tailored for structural-dynamics surrogate modeling, where the motivation and evaluated signal behind this design are illustrated in Figure 1. The model builds graph resolutions for irregular structures, recomputes edge travel times at each resolution, encodes only historical load samples, and applies differentiable arrival gates before P-wave, S-wave, and mixed message branches update node states. This design aims to keep the efficiency of multiresolution graph learning while preserving a conservative physical constraint. Coarse edges may reduce graph distance, but they do not erase travel time.

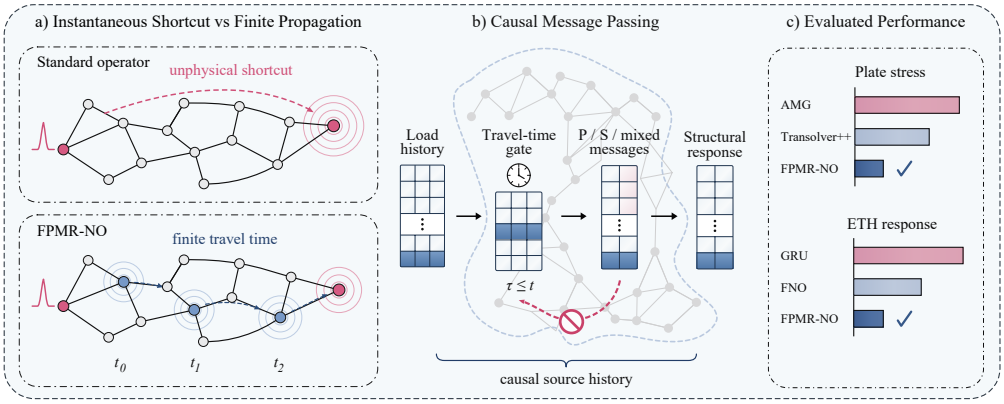

Figure 1: Introductory overview. Standard graph-based operators can create instantaneous computational shortcuts because graph connectivity does not by itself encode physical arrival order. FPMR-NO turns this contrast into an architectural prior by gating messages with finite travel time and causal source history, and the evaluated stress and sparse impact-response tasks show the intended performance signal.

We validate FPMR-NO in three structural-dynamics settings. On DeformingPlate, FPMR-NO predicts one-step stress residuals on variable-size tetrahedral meshes, testing finite-propagation-aware multiresolution learning for irregular-mesh stress modeling. On the ETH four-story laboratory structure, FPMR-NO builds a cross-spectral initialization from training data and combines it with a causal modal response head to map impact histories to sparse acceleration responses. This setting shows how the same operator family handles sensor-based, frequency-sensitive impact-response transfer from a real structural testbed. Finally, we construct controlled graph-wave and delayed-response experiments that use known arrival times, wave-speed counterfactuals, and future-input interventions to directly observe the effects of finite-propagation gating (PG) and causal load encoding (CLE).

Our contributions are listed below. 1) We introduce FPMR-NO, which recomputes material travel times at each graph scale and gates message passing by arrival conditions, so coarse aggregation shortens computational paths without erasing propagation time. 2) We design causal load-history encoding with P-wave, S-wave, and mixed message branches that jointly model source history, material wave speed, and edge arrival conditions to suppress instantaneous graph shortcuts in dynamic response prediction. 3) We obtain the lowest normalized relative L2 on DeformingPlate, 0.0260, improving over AMG by 2.4% and Transolver by 1.7%; on ETH, FPMR-NO reduces macro-channel NRMSE to 0.86, improving over the best baseline by 14.1%; in controlled graph-wave experiments, it reduces relative L2 by 71.8% and pre-arrival energy by 13.3×.

2 Related Work

Neural operators. DeepONet established a branch–trunk route for learning nonlinear operators (Lu et al., 2021), while the general neural-operator formulation and Fourier neural operator (FNO)

error analysis formalize maps between function spaces (Kovachki et al., 2023; 2021). FNOs learn such mappings by parameterizing kernels in spectral space (Li et al., 2021), and subsequent spectral and latent variants improve efficiency, resolution handling, or local-global mixing (Tran et al., 2023; Wu et al., 2023). Poseidon extends operator learning toward pretrained PDE foundation models (Herde et al., 2024), and Riesz Neural Operator (RNO) uses spectral derivative structure to blend local and global information (Liu et al., 2026). These directions motivate stronger inductive biases, but they do not directly encode finite travel time on structural graphs.

Geometry-aware and graph-based operators. MeshGraphNets learn simulators directly on mesh graphs and have demonstrated strong performance on structural and fluid systems (Pfaff et al., 2021). Graph and mesh neural operators further target non-Cartesian domains through multipole graph kernels, message passing, and mesh-agnostic solvers (Li et al., 2020; Brandstetter et al., 2022; Boussif et al., 2022). Recent arbitrary-geometry operators include learned-deformation FNOs and GINO for complex 3D geometries (Li et al., 2023a;b); AMG, a multi-graph framework for PDEs on arbitrary geometries (Li et al., 2025); Transolver and Transolver++, which learn physical-state slices for efficient PDE solving on complex geometries (Wu et al., 2024; Luo et al., 2025); GNOT, NUNO, and LRQ-Solver, which use transformer or non-uniform operator-learning designs for general or large-scale 3D PDEs (Hao et al., 2023; Liu et al., 2023; Zeng et al., 2026); RIGNO, which uses regional interaction graphs for arbitrary domains (Mousavi et al., 2025); and lightweight convolutional operators such as D-SENO (Chauhan et al., 2026). FPMR-NO is closest to this multiscale and graph-based line of work, but it changes the role of graph edges. An edge provides a possible computational route only after the corresponding source-to-destination arrival condition is satisfied.

Structural response and wave-propagation priors. Structural health monitoring and vibration-based damage detection have long treated measured structural response as a sensing-driven inference problem (Farrar & Worden, 2010; Avci et al., 2021). Recent physics-guided deep surrogates for seismic response add equation-level constraints to neural predictors (Zhang et al., 2020), while wave-equation physics-informed neural network (PINN) studies show how explicit wave physics can improve extrapolation beyond observed boundary data (Moseley et al., 2020). Closely related operator-learning studies interface neural operators with finite elements for multiscale mechanics, predict 3D elastic wave propagation with factorized FNOs, and estimate structural-dynamics parameters through neural-operator surrogates (Yin et al., 2022; Lehmann et al., 2024; Zhou et al., 2025). These works motivate physics-aware learning for dynamic structures, but their constraints are usually expressed through losses, prescribed sensor/time-series models, or continuous coordinates.

3 METHOD

The overall framework of our method is shown in Figure 2. At a high level, FPMR-NO first constructs multiresolution structural graphs and encodes the observed load history into node representations. Material-aware travel times then determine when P-wave, S-wave, and mixed messages can propagate at each graph scale, after which the resulting multiscale features are mapped back to the fine resolution, fused, and decoded for the target structural response. Appendix A provides further details on travel-time construction, gated message passing, and causal encoding in Figures 7–9.

3.1 PROBLEM SETUP

Let a structural state be observed on nodes V with coordinates $\boldsymbol{x}_i$, static features $\boldsymbol{s}_i$, dynamic state $\boldsymbol{u}_i(t)$, and load history $\ell_i(1{:}t)$. Depending on the task, the model predicts a next state, residual, acceleration, stress, or sensor response. We write the graph at resolution ℓ as $\mathcal{G}^{(\ell)} = (V^{(\ell)}, E^{(\ell)})$ and the mapping from fine node i to its node at scale ℓ as $\phi_\ell(i)$. The first scale is the original graph; coarser scales are deterministic geometric regions or task-defined clusters. Throughout, bold lowercase symbols denote column vectors, ordinary italic lowercase symbols denote scalars, and bold uppercase symbols denote matrices. Thus $\boldsymbol{x}_i$, $\boldsymbol{s}_i$, $\boldsymbol{u}_i(t)$, $\boldsymbol{h}_i^\ell$, $\boldsymbol{e}_{ij}$, $\boldsymbol{m}_{ij}$, and $\boldsymbol{y}_i$ are vectors, whereas $\ell_i(s)$, d_{ij}, c_i^m, τ_{ij}^m, a_i^m, Δ_{ij}^m, g_{ij}^m, and η are scalars. The symbols f_b^r, U_r, F, and D denote vector-valued maps rather than matrices.

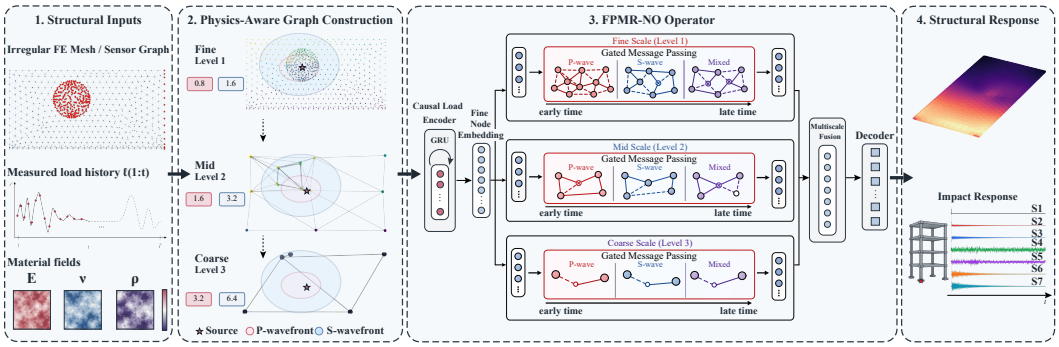

Figure 2: Overall framework of FPMR-NO. Structural inputs are encoded into multiresolution graphs, where causal load history and material-aware travel-time gates control P-wave, S-wave, and mixed message passing. The fused multiscale representation is decoded into structural stress or sparse impact-response predictions.

3.2 Material-Aware Travel-Time Graphs

For an edge $i \to j$, the operator computes the edge length $d_{ij} = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2$ and P/S travel times by harmonic node-speed averaging,

$$\tau_{ij}^{p} = \frac{d_{ij}}{2}\left(\frac{1}{c_i^{p}} + \frac{1}{c_j^{p}}\right), \qquad \tau_{ij}^{s} = \frac{d_{ij}}{2}\left(\frac{1}{c_i^{s}} + \frac{1}{c_j^{s}}\right). \tag{1}$$

When elastic constants are available, c^p and c^s are computed from Lamé parameters and density. At every graph scale, coordinates and wave speeds are re-averaged inside clusters before edge travel times are recomputed. The model then runs a shortest-path arrival calculation from the source set S.

$$a_i^{m} = \min_{\pi: S \rightsquigarrow i} \sum_{(u,v)\in\pi} \tau_{uv}^{m}, \qquad m \in \{p, s\}. \tag{2}$$

This step gives multiresolution learning its physical interpretation. A coarse graph edge may reduce computational depth, but cross-region communication remains governed by recomputed travel times rather than acting as an unconstrained zero-time shortcut, as summarized in Appendix Figure 7.

3.3 Causal Load Encoding

FPMR-NO encodes only the load history available up to the current query time. A gated recurrent unit (GRU) maps the node-major sequence $\ell_i(1{:}t)$ to a hidden context $\boldsymbol{h}_i^{\ell}$. A multilayer perceptron (MLP) then produces the initial node state.

$$\boldsymbol{h}_i^{(0)} = \mathrm{MLP}_{\mathrm{enc}}\left([\boldsymbol{s}_i, \boldsymbol{u}_i(t), \boldsymbol{h}_i^{\ell}]\right). \tag{3}$$

The w/o CLE ablation replaces this context by zeros. Causal variants never receive future load samples; future-intervention tests therefore measure whether a prediction changes when only future samples are shuffled. Appendix Figure 9 summarizes the causal-history path and its diagnostics.

3.4 Finite-Propagation Message Passing

For each scale and each edge $i \to j$, FPMR-NO constructs a message input from source state, destination state, edge features, and the two travel times. The availability margin and differentiable gate for mode m at query time t are

$$\Delta_{ij}^{m}(t) = t - a_i^{m} - \tau_{ij}^{m}. \tag{4}$$

$$g_{ij}^{m}(t) = \begin{cases} \mathbf{1}[\Delta_{ij}^{m}(t) \geq 0], & \eta \leq 0, \\ \sigma(\Delta_{ij}^{m}(t)/\eta), & \eta > 0, \end{cases} \tag{5}$$

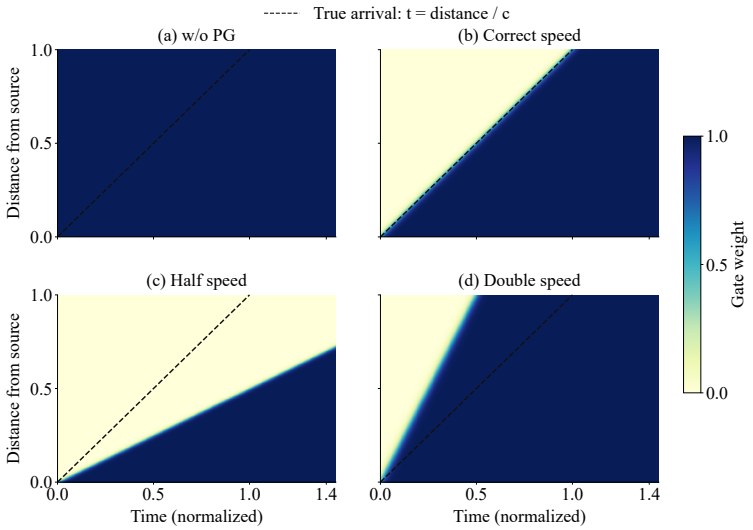

Figure 3: Analytical illustration of the travel-time gate in Eq. equation 5. Colors show gate weights for a source at position zero, normalized true speed $c = 1$, and softness $\eta = 0.0125$; dashed lines mark the true arrival $t = d/c$. The w/o PG variant leaves all connections open. The correct-speed gate follows the arrival line, while half and double speeds delay and advance opening.

where η is a softness parameter. For $\eta \leq 0$, the gate enforces exact arrival support and suppresses all pre-arrival messages. For $\eta > 0$, it provides a smooth approximation in which pre-arrival communication is exponentially attenuated rather than identically zero. Figure 3 illustrates how the gate opens around the arrival boundary and shifts when the assumed wave speed changes. Separate P-wave, S-wave, and mixed branches produce

$$\boldsymbol{m}^p_{ij}(t) = g^p_{ij}(t)\, f_p([\boldsymbol{h}_i(t), \boldsymbol{h}_j(t), \boldsymbol{e}_{ij}, \tau^p_{ij}, \tau^s_{ij}]), \tag{6}$$

$$\boldsymbol{m}^s_{ij}(t) = g^s_{ij}(t)\, f_s([\boldsymbol{h}_i(t), \boldsymbol{h}_j(t), \boldsymbol{e}_{ij}, \tau^p_{ij}, \tau^s_{ij}]), \tag{7}$$

$$\boldsymbol{m}^x_{ij}(t) = \max(g^p_{ij}(t), g^s_{ij}(t))\, f_x([\boldsymbol{h}_i(t), \boldsymbol{h}_j(t), \boldsymbol{e}_{ij}, \tau^p_{ij}, \tau^s_{ij}]). \tag{8}$$

Mean-aggregated branch messages update each node through a residual MLP and layer normalization. Each graph scale performs a predetermined number of propagation steps. The resulting scale states are mapped back to the fine graph, concatenated, fused by an MLP, and decoded to the task output. Appendix Figure 8 depicts the three gated message branches.

Assumption 1 (Finite-propagation access)**.** *Let a query time $t \geq 0$ be given. At every graph scale, FPMR-NO satisfies the following access rules.*

1. *The load encoder at node i receives only $\ell_i(s)$ for $s \leq t$ and never receives future load samples $\ell_i(s)$ for $s > t$.*
2. *Edge travel times are nonnegative, and source-to-node arrival values are computed as shortest-path sums of those travel times at the same graph scale.*
3. *Each P/S branch message is multiplied by its arrival gate in Eq. equation 5, and the mixed branch is multiplied by $\max(g^p, g^s)$.*
4. *Node updates aggregate only current node states and incoming gated messages. Restriction, prolongation, fusion, and decoding are deterministic functions of their inputs and do not access future load samples. Scale-transfer maps are local cell maps. They move information between a coarse cell and the fine nodes represented by that cell, but do not connect unrelated cells except through graph message passing.*

Assumption 1 is stated at the resolution where each state lives. When a coarse cell state is copied back to all fine nodes inside that cell, the certified spatial unit is the coarse cell representation;

Table 1: DeformingPlate stress prediction under the official AMG one-step setting. Rows follow chronological method order, with FPMR-NO last. Values are mean ± population standard deviation over three seeds; standard deviations are shown to four decimals, and bold indicates the best mean in each performance column.

Method	Params	Lat. ms	Norm. rel. L2 ↓	Phys. rel. L2 ↓
MLP	0.07M	1.06	3.18 ± 0.0133	3.15 ± 0.0179
MeshGraphNet	2.01M	10.07	2.76 ± 0.0064	2.72 ± 0.0354
GNOT	1.39M	2.17	24.11 ± 0.1969	23.78 ± 0.2795
Transolver	0.71M	11.72	2.65 ± 0.0716	2.59 ± 0.0016
AMG	2.38M	4.40	2.66 ± 0.0588	2.62 ± 0.0322
RIGNO	0.64M	4.32	2.78 ± 0.0541	2.67 ± 0.0099
Transolver++	0.44M	10.09	2.74 ± 0.0880	2.59 ± 0.0289
FPMR-NO	0.21M	9.70	$\mathbf{2.60 \pm 0.0483}$	$\mathbf{2.58 \pm 0.0146}$

cross-cell graph communication remains time-gated. Appendix A.4 gives the complete proof of the finite-propagation and causal-history access properties induced by this assumption.

3.5 Task Adapters

The general operator is specialized only where the task output requires it. On DeformingPlate, FPMR-NO predicts the one-step stress residual on variable-size tetrahedral meshes using the official AMG stress task; Figure 12 visualizes the resulting multiresolution graph. Because the released metadata lacks material constants and a valid time step, the formal Plate model uses PG as a structural graph bias rather than as strict P/S arrival identification. On ETH impact events, FPMR-NO uses a causal modal-propagation head. Training-only cross spectra provide transfer initialization (TI) for a truncated impulse-response basis, and a geometry residual (GR) conditions coefficients on sensor and impact metadata. This sparse-sensor adapter predicts measured accelerometer channels rather than full-field displacement or stress.

4 Experiments

Implementation. We use task-specific inputs and prediction heads with common training and evaluation procedures. Formal runs use seeds 2024, 2025, and 2026, validation-only hyperparameter and model selection, and held-out test evaluation after selection. The Plate operator uses two graph scales and two message-passing steps, while the ETH head uses 32 modes. Unless noted otherwise, lower-is-better errors in the experimental analysis are coefficients of 10^{-2}; percentages, latency, and memory retain their stated units. We report mean and population standard deviation across seeds. Timing and peak graphics processing unit (GPU) memory are measured on an NVIDIA A40.

Datasets. For DeformingPlate, we follow the official AMG one-step stress setting, retain the 1200/100/100 train-validation-test trajectory split, construct variable-size tetrahedral graphs, and evaluate the one-step stress residual after inverse normalization. For ETH, we extract eight impact events from each of 32 records, standardize hammer and accelerometer channels using training statistics, and form grouped, direction-stratified splits with 192 training, 32 validation, and 32 test events. Controlled mechanism experiments use synthetic graph-wave and delayed-response systems with known arrival times and future-intervention targets.

Baselines. We use MLP, MeshGraphNet (Pfaff et al., 2021), GNOT (Hao et al., 2023), Transolver (Wu et al., 2024), AMG (Li et al., 2025), RIGNO (Mousavi et al., 2025), and Transolver++ (Luo et al., 2025) as baselines on DeformingPlate. On ETH, we use MLP, long short-term memory (LSTM), Transformer, DeepONet (Lu et al., 2021), Poseidon (Herde et al., 2024), and RNO (Liu et al., 2026) as baselines.

4.1 Main Results

Table 2: ETH sparse impact-response transfer. Rows follow chronological method order, with FPMR-NO last. Error columns use the section-wide reporting scale. Standard deviations are shown to four decimals; bold indicates the best mean in each performance column.

Method	Params	Lat. ms	Macro NRMSE ↓	Spectral rel. L2 ↓
MLP	7.44M	4.76	100.56 ± 1.5435	68.89 ± 1.4744
LSTM	0.06M	7.71	102.74 ± 1.3156	86.18 ± 3.2122
Transformer	0.53M	7.01	100.34 ± 1.0618	72.03 ± 1.2937
DeepONet	0.56M	3.06	100.01 ± 0.0033	99.99 ± 0.0015
Poseidon	0.11M	5.33	104.57 ± 4.6431	94.16 ± 4.9309
RNO	0.07M	5.93	105.29 ± 2.1229	101.84 ± 4.2516
FPMR-NO	0.14M	3.26	$\mathbf{85.92 \pm 0.0551}$	$\mathbf{56.92 \pm 0.0770}$

Table 1 shows the lowest mean normalized relative L2 for FPMR-NO, 2.60, improving over AMG by 2.37% and Transolver by 1.67%. Figures 10 and 13 complement these averages with nodal and surface views of the same test case. Under the shared stress scale, the prediction preserves the location and overall structure of the high-stress region, supporting the qualitative agreement indicated by the scalar metrics.

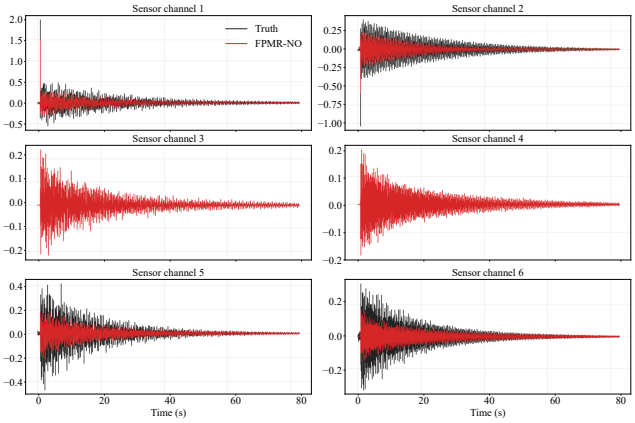

Figure 4: Representative ETH impact response. Measured and predicted acceleration histories are overlaid for the same test event across all six response channels.

On ETH, FPMR-NO achieves the lowest macro-channel NRMSE and spectral relative L2 (Table 2). Macro-NRMSE improvements are 14.1% over DeepONet, 14.4% over Transformer, and 18.4% over RNO. The event histories in Figure 4 are complemented by the spectra in Figure 14. Dominant peak locations are reproduced, while several peak amplitudes and the low-frequency background remain underestimated. Thus the frequency-domain advantage coexists with residual amplitude mismatch.

4.2 Mechanism Evidence

Controlled synthetic systems expose arrival times and load histories that Plate and ETH do not jointly identify. Table 3 and Figure 11 separate finite-propagation and causal-history effects. On graph waves with known arrivals, enabling the gate reduces relative L2 from 4.42 to 1.25, pre-arrival energy from 0.0836 to 0.0063, and arrival mean absolute error (MAE) from 1.13 to 1.02. The 13.3-fold energy reduction directly measures suppression of response before physical arrival. Increasing causal history from 8 to 32 steps reduces NRMSE from 52.69 to 7.97 while retaining exact zero prediction change under future-only interventions, since future samples are excluded. The noncausal oracle reaches NRMSE 6.51, but shuffling future inputs changes its prediction by 20.62. Its additional accuracy therefore relies on unavailable future information, whereas the longer causal history improves prediction without that access.

Table 3: Controlled mechanism results. Separate panels report finite-propagation and causal-history diagnostics under identifiable synthetic conditions. Exact-zero future change denotes invariance to future-only interventions; otherwise standard deviations are shown to four decimals, and bold indicates the best mean in each comparable column.

Finite-propagation diagnostic

Method	Rel. L2 ↓	Pre-arrival energy ↓	Arrival MAE ↓
FPMR-NO	$\mathbf{1.25 \pm 0.0871}$	$\mathbf{0.0063 \pm 0.0013}$	$\mathbf{1.02 \pm <0.0001}$
w/o PG	4.42 ± 0.5944	0.0836 ± 0.0273	1.13 ± 0.0325

Causal-history diagnostic

Variant	NRMSE ↓	Future change ↓
history 8	52.69 ± 0.1410	**0** (exact)
history 32	7.97 ± 0.3183	**0** (exact)
noncausal oracle	$\mathbf{6.51 \pm 0.2783}$	20.62 ± 0.2493

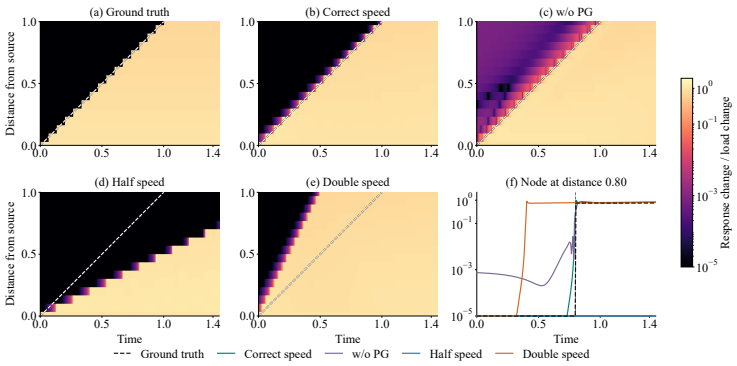

Figure 5: Source-load influence in a single-scale graph-wave experiment. Colors show $I_j(t) = |\hat{y}_j(t; 1.2) - \hat{y}_j(t; 1.0)|/0.2$ on a shared logarithmic scale, with display floor 10^{-5}. Dashed lines mark the true arrival $t = d/c$, where $c = 1$. The correct-, half-, and double-speed panels use the same trained model; the w/o PG model is trained separately. The final panel compares influence at distance 0.80. These are prediction sensitivities, not attention weights or analytical gate values.

Figure 5 directly visualizes how a source-load perturbation changes predicted responses in a separate, same-seed, single-scale graph-wave experiment. With the correct speed, pre-arrival influence concentrates near the arrival boundary; without gating, weaker early influence spreads over more distant node–time pairs. Halving or doubling the assumed gate speed delays or advances the response. Localization does not imply lower total sensitivity. The pre-arrival influence root mean square (RMS) is 6.69 for the soft gate and 4.01 without gating in this case. Appendix B.3 details the setup and this distinction.

4.3 Ablations, Robustness, and Efficiency

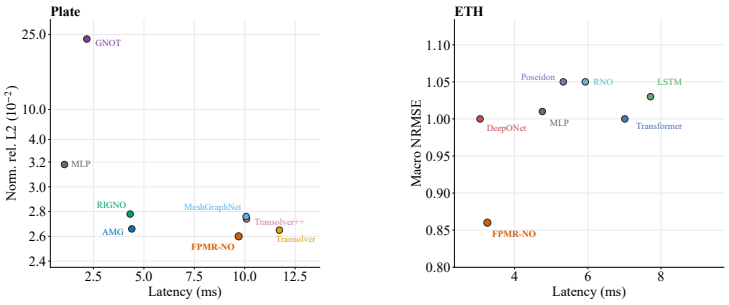

Figure 6: Accuracy-efficiency trade-off under the common evaluation setup. The plot compares mean test error with observed inference latency for the evaluated methods.

The ablations are task-specific. Plate directly evaluates multiresolution aggregation (MR), propagation gating (PG), and causal load encoding (CLE), while ETH evaluates the task-specific transfer initialization (TI) and geometry residual (GR) used for sparse sensor-response prediction. The effects of PG and CLE are further isolated in the controlled mechanism experiments. Table 4 identifies the useful components, while Figure 6 shows the accuracy-efficiency trade-off. On Plate, w/o MR and w/o PG increase error by 6.76% and 3.01%, respectively. w/o CLE increases the primary error to 2.75 ± 0.0347, a 5.94% increase over FPMR-NO, and gives a physical relative L2 of 2.63 ± 0.0498. Figure 12 illustrates the spatial aggregation. 1305 fine nodes are represented by 17 coarse regions. On ETH, w/o TI increases macro NRMSE by 8.15%. w/o GR improves it by 0.47% but worsens spectral relative L2 from 56.92 to 57.54. TI drives the larger gain, while GR trades time-domain and spectral accuracy.

Figure 15 shows distinct responses to input perturbations. ETH macro-NRMSE degradation stays below 6.6% even with 50% missing channels. Plate tolerates small geometry jitter but degrades sharply with noise and missing nodes, indicating greater dependence on dense spatial observations.

Table 6 tests reduced training data. At 10%, 25%, and 50% fractions, FPMR-NO improves over Transolver on Plate by 4.93%, 1.39%, and 0.81%, respectively, and over DeepONet on ETH by 10.59%, 10.46%, and 12.31%. The Plate margin is largest at the smallest training fraction. ETH retains a larger mean advantage but has higher seed variability at 10%.

Table 4: Task-specific FPMR-NO ablations. Values are test metrics after the ablation configuration is selected. Error metrics use the section-wide reporting scale, standard deviations are shown to four decimals, and bold indicates the best mean within each dataset and metric column.

Dataset	Method	Primary metric ↓	Secondary metric	Degradation
Plate	FPMR-NO	$\mathbf{2.60 \pm 0.0483}$	phys. L2 $\mathbf{2.58 \pm 0.0146}$	0.00%
Plate	w/o MR	2.78 ± 0.0142	phys. L2 2.74 ± 0.0193	6.76%
Plate	w/o PG	2.68 ± 0.0763	phys. L2 2.59 ± 0.0229	3.01%
Plate	w/o CLE	2.75 ± 0.0347	phys. L2 2.63 ± 0.0498	5.94%
ETH	FPMR-NO	85.92 ± 0.0551	spectral L2 $\mathbf{56.92 \pm 0.0770}$	0.00%
ETH	w/o TI	92.92 ± 0.3240	spectral L2 62.18 ± 0.3180	8.15%
ETH	w/o GR	$\mathbf{85.52 \pm 0.0581}$	spectral L2 57.54 ± 0.0367	−0.47%
ETH	w/o TI + GR	92.44 ± 0.7090	spectral L2 63.22 ± 0.3400	7.59%

Figure 16 probes temporal behavior with selected model states. Plate errors accumulate during autoregressive rollout and flatten at a high level, revealing a gap between one-step accuracy and long-horizon stability. ETH error on training-standardized response prefixes decreases as the window grows from 256 to 4096 samples. This evaluates windowed reconstruction rather than recursive extrapolation.

The accuracy–latency panel in Figure 6 places FPMR-NO on a trade-off. Plate accuracy improves over AMG at higher latency, while ETH combines its lowest error with 3.26 ms/event, 0.14M parameters, and 0.07 GB peak memory. Plate also uses fewer parameters and less peak memory than AMG and MeshGraphNet. In the separate synchronized, batch-size-one measurements of Figure 17, Plate stays below 5 ms up to 4096 synthetic nodes, and ETH remains near 0.8 ms across sequence lengths 512–8192. Appendix B.6 reports validation-only one-factor sensitivity, and Appendix B.7 reports paired t-test comparisons on ETH with Holm correction.

5 Discussion

The results suggest that FPMR-NO is most useful when structural response prediction requires both irregular-geometry reasoning and temporally causal information flow. Its advantage comes from aligning the operator with structural-dynamics constraints rather than from increasing model size. MR supplies geometric context, PG suppresses messages before identifiable arrivals, and CLE blocks future inputs. This also explains the task split. DeformingPlate yields a close mean advantage over strong geometry operators, while ETH benefits more directly from causal impact history and TI. The current limits are task-specific. Plate lacks valid time steps, material constants, explicit loads, and wave speeds, so PG acts as a structural graph bias; ETH provides sparse accelerometer channels rather than full-field displacement or stress.

6 Conclusion

We presented FPMR-NO, combining CLE, material travel-time gates, and MR graph propagation. Its central idea is to condition spatial message exchange on propagation time while restricting load inputs to the observed past. Under the stated assumptions, the analysis establishes future-load invariance and characterizes hard-gate support and soft-gate leakage. Controlled experiments show reduced pre-arrival energy and invariance to future-only interventions. On DeformingPlate, FPMR-NO achieves the lowest mean normalized relative L2 among the evaluated methods, improving over AMG and Transolver by 2.37% and 1.67%, respectively. On ETH, it reduces macro NRMSE by 14.1% relative to DeepONet and attains the lowest spectral relative L2. Ablations identify the contributions of MR and TI. Together, these findings support propagation-aware information flow as a useful inductive bias for structural-response prediction on irregular meshes and at sparse sensors. Future work will extend FPMR-NO to structural-dynamics datasets with explicit material parameters, time-step information, and richer full-field measurements, enabling more direct validation of physically calibrated P/S travel-time gating. We also plan to investigate adaptive multiresolution graph construction for longer-horizon and more heterogeneous structural systems.

Ethics Statement

This work studies surrogate modeling for structural dynamics and structural-health-monitoring-style response prediction. The reported ETH dataset is a laboratory-scale structural testbed rather than human-subject data. Potential downstream risk arises if a learned surrogate is used as a substitute for validated engineering analysis in safety-critical decisions. The method should therefore be treated as a research surrogate and must be calibrated, verified, and bounded before deployment in structural design, inspection, or emergency assessment.

Reproducibility Statement

We release an anonymized GitHub repository with code and scripts for end-to-end reproduction of all experiments and parameters. The link is available at: https://anonymous.4open.science/r/FPMR-NO-6FCF

AI Use Statement

ChatGPT was used only for language polishing, including sentence rephrasing, grammar checking, readability improvement, and flow refinement. It was not involved in idea generation, methodology, experimental design, data generation, result calculation, hyperparameter selection, or scientific analysis. All concepts, experiments, analyses, and conclusions were developed, conducted, and verified by the authors, who take full responsibility for the manuscript. In addition, ChatGPT was used to assist in writing Python code for some experimental visualizations. The authors reviewed and verified the code and visualizations.

References

Onur Avci, Osama Abdeljaber, Serkan Kiranyaz, Mohammed Hussein, Moncef Gabbouj, and Daniel J. Inman. A review of vibration-based damage detection in civil structures: From traditional methods to machine learning and deep learning applications. *Mechanical Systems and Signal Processing*, 147:107077, 2021. ISSN 0888-3270. doi: 10.1016/j.ymssp.2020.107077. URL https://www.sciencedirect.com/science/article/pii/S0888327020304635.

Oussama Boussif, Yoshua Bengio, Loubna Benabbou, and Dan Assouline. Magnet: Mesh agnostic neural PDE solver. In Sanmi Koyejo, S. Mohamed, A. Agarwal, Danielle Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022. URL http://papers.nips.cc/paper_files/paper/2022/hash/cf4c7ee0734cdfe09a099cf6cd7b117a-Abstract-Conference.html.

Matteo Brambilla, Vasilis Dertimanis, Eleni Chatzi, Paolo Chiariotti, and Alfredo Cigada. Experimental campaign for dynamic characterization of a laboratory-scale structure, 2025. URL https://doi.org/10.5281/zenodo.15516419. Dataset.

Johannes Brandstetter, Daniel E. Worrall, and Max Welling. Message passing neural PDE solvers. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022. URL https://openreview.net/forum?id=vSix3HPYKSU.

Prajwal Chauhan, Salah Eddine Choutri, and Saif Jabari. Efficient dilated squeeze and excitation neural operator for differential equations. *Transactions on Machine Learning Research*, 2026. ISSN 2835-8856. URL https://openreview.net/forum?id=Xl942THEUa.

Charles R. Farrar and Keith Worden. An introduction to structural health monitoring. In Arnaud Deraemaeker and Keith Worden (eds.), *New Trends in Vibration Based Structural Health Monitoring*, pp. 1–17. Springer Vienna, Vienna, 2010. ISBN 978-3-7091-0399-9. doi: 10.1007/978-3-7091-0399-9_1. URL https://doi.org/10.1007/978-3-7091-0399-9_1.

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 Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, pp. 12556–12569. PMLR, 2023. URL https://proceedings.mlr.press/v202/hao23c.html.

Maximilian Herde, Bogdan Raonic, Tobias Rohner, Roger Kappeli, Roberto Molinaro, Emmanuel de Bezenac, and Siddhartha Mishra. Poseidon: Efficient foundation models for PDEs. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 72525–72624. Curran Associates, Inc., 2024. doi: 10.52202/079017-2311. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/84e1b1ec17bb11c57234e96433022a9a-Paper-Conference.pdf.

Sepp Hochreiter and Juergen Schmidhuber. Long short-term memory. *Neural Computation*, 9 (8):1735–1780, 1997. doi: 10.1162/neco.1997.9.8.1735. URL https://dblp.org/rec/journals/neco/HochreiterS97.bib.

Nikola B. Kovachki, Samuel Lanthaler, and Siddhartha Mishra. On universal approximation and error bounds for fourier neural operators. *J. Mach. Learn. Res.*, 22:290:1–290:76, 2021. URL https://jmlr.org/papers/v22/21-0806.html.

Nikola B. Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew M. Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to pdes. *J. Mach. Learn. Res.*, 24:89:1–89:97, 2023. URL https://jmlr.org/papers/v24/21-1524.html.

Fanny Lehmann, Filippo Gatti, Michaël Bertin, and Didier Clouteau. 3d elastic wave propagation with a factorized fourier neural operator (F-FNO). *Computer Methods in Applied Mechanics and Engineering*, 420:116718, 2024. ISSN 0045-7825. doi: 10.1016/j.cma.2023.116718. URL https://www.sciencedirect.com/science/article/pii/S0045782523008411.

Zhihao Li, Haoze Song, Di Xiao, Zhilu Lai, and Wei Wang. Harnessing scale and physics: A multi-graph neural operator framework for PDEs on arbitrary geometries. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.1*, KDD '25, pp. 729–740, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400712456. doi: 10.1145/3690624.3709173. URL https://doi.org/10.1145/3690624.3709173.

Zongyi Li, Nikola B. Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Andrew M. Stuart, Kaushik Bhattacharya, and Anima Anandkumar. Multipole graph neural operator for parametric partial differential equations. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin (eds.), *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/4b21cf96d4cf612f239a6c322b10c8fe-Abstract.html.

Zongyi Li, Nikola Borislavov Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=c8P9NQVtmnO.

Zongyi Li, Daniel Zhengyu Huang, Burigede Liu, and Anima Anandkumar. Fourier neural operator with learned deformations for pdes on general geometries. *J. Mach. Learn. Res.*, 24:388:1–388:26, 2023a. URL http://jmlr.org/papers/v24/23-0064.html.

Zongyi Li, Nikola B. Kovachki, Christopher B. Choy, Boyi Li, Jean Kossaifi, Shourya Prakash Otta, Mohammad Amin Nabian, Maximilian Stadler, Christian Hundt, Kamyar Azizzadenesheli, and Animashree Anandkumar. Geometry-informed neural operator for large-scale 3d pdes. In Alice

Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine (eds.), *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023b. URL http://papers.nips.cc/paper_files/paper/2023/hash/70518ea42831f02afc3a2828993935ad-Abstract-Conference.html.

Shouyi Liu, Xiaokang Yang, and Yuntian Chen. Riesz neural operator for solving partial differential equations. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=Vjw7q1quNt.

Songming Liu, Zhongkai Hao, Chengyang Ying, Hang Su, Ze Cheng, and Jun Zhu. NUNO: A general framework for learning parametric pdes with non-uniform data. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, pp. 21658–21671. PMLR, 2023. URL https://proceedings.mlr.press/v202/liu23o.html.

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. *Nat. Mach. Intell.*, 3(3):218–229, 2021. doi: 10.1038/S42256-021-00302-5. URL https://doi.org/10.1038/s42256-021-00302-5.

Huakun Luo, Haixu Wu, Hang Zhou, Lanxiang Xing, Yichen Di, Jianmin Wang, and Mingsheng Long. Transolver++: An accurate neural solver for PDEs on million-scale geometries. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 41432–41449. PMLR, 13–19 Jul 2025. URL https://proceedings.mlr.press/v267/luo25o.html.

Ben Moseley, Andrew Markham, and Tarje Nissen-Meyer. Solving the wave equation with physics-informed deep learning, 2020. URL https://arxiv.org/abs/2006.11894.

Sepehr Mousavi, Shizheng Wen, Levi Lingsch, Maximilian Herde, Bogdan Raonic, and Siddhartha Mishra. RIGNO: A graph-based framework for robust and accurate operator learning for PDEs on arbitrary domains. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, pp. 150039–150101. Curran Associates, Inc., 2025. doi: 10.52202/085713-5018. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/dcb91f43033bb1d367d1848806dee98d-Paper-Conference.pdf.

Tobias Pfaff, Meire Fortunato, Alvaro Sanchez-Gonzalez, and Peter W. Battaglia. Learning mesh-based simulation with graph networks. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=roNqYL0_XP.

Alasdair Tran, Alexander Patrick Mathews, Lexing Xie, and Cheng Soon Ong. Factorized fourier neural operators. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023. URL https://openreview.net/forum?id=tmIiMPl4IPa.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems 30*, pp. 5998–6008, 2017. URL https://dblp.org/rec/conf/nips/VaswaniSPUJGKP17.bib.

Haixu Wu, Tengge Hu, Huakun Luo, Jianmin Wang, and Mingsheng Long. Solving high-dimensional pdes with latent spectral models. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, pp. 37417–37438. PMLR, 2023. URL https://proceedings.mlr.press/v202/wu23f.html.

Haixu Wu, Huakun Luo, Haowen Wang, Jianmin Wang, and Mingsheng Long. Transolver: A fast transformer solver for PDEs on general geometries. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 53681–53705. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/wu24r.html.

Minglang Yin, Enrui Zhang, Yue Yu, and George Em Karniadakis. Interfacing finite elements with deep neural operators for fast multiscale modeling of mechanics problems. *Computer Methods in Applied Mechanics and Engineering*, 402:115027, 2022. ISSN 0045-7825. doi: 10.1016/j.cma.2022.115027. URL https://www.sciencedirect.com/science/article/pii/S0045782522002535. A Special Issue in Honor of the Lifetime Achievements of J. Tinsley Oden.

Peijian Zeng, Guan Wang, Haohao Gu, Xiaoguang Hu, Tiezhu Gao, Zhuowei Wang, Aimin Yang, and Xiaoyu Song. LRQ-Solver: A transformer-based neural operator for fast and accurate solving of large-scale 3d pdes. *Computer-Aided Design*, 200:104151, 2026. ISSN 0010-4485. doi: 10.1016/j.cad.2026.104151. URL https://www.sciencedirect.com/science/article/pii/S0010448526001211.

Ruiyang Zhang, Yang Liu, and Hao Sun. Physics-guided convolutional neural network (PhyCNN) for data-driven seismic response modeling. *Engineering Structures*, 215:110704, 2020. ISSN 0141-0296. doi: 10.1016/j.engstruct.2020.110704. URL https://www.sciencedirect.com/science/article/pii/S0141029619345080.

Mingyuan Zhou, Haoze Song, Wenjing Ye, Wei Wang, and Zhilu Lai. Parameter estimation of structural dynamics with neural operators enabled surrogate modeling. *Mechanical Systems and Signal Processing*, 237:112914, 2025. ISSN 0888-3270. doi: 10.1016/j.ymssp.2025.112914. URL https://www.sciencedirect.com/science/article/pii/S0888327025006156.

A METHOD DETAILS AND PROOF

Figures 7–9 expand the three mechanisms in Figure 2. We use r for graph resolution, k for message-passing iteration, and $m \in \{p, s\}$ for propagation mode.

A.1 MATERIAL-AWARE TRAVEL-TIME CONSTRUCTION

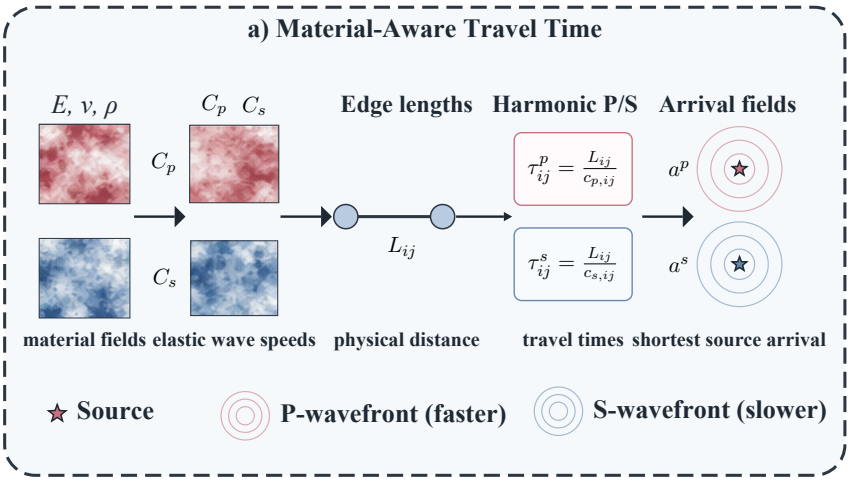

Figure 7: Material-aware travel-time construction. Material fields determine P/S speeds, which combine with edge lengths to define edge travel times. Shortest-path accumulation from the source set produces mode-specific arrival fields at each graph resolution.

Figure 7 connects local material properties to the arrival fields used by the gates. For isotropic linear elasticity, Young's modulus $E_i > 0$, Poisson's ratio $-1 < \nu_i < 1/2$, and density $\rho_i > 0$ give

$$\lambda_i = \frac{E_i \nu_i}{(1+\nu_i)(1-2\nu_i)}, \qquad \mu_i = \frac{E_i}{2(1+\nu_i)}, \tag{9}$$

$$c_i^p = \sqrt{\frac{\lambda_i + 2\mu_i}{\rho_i}}, \qquad c_i^s = \sqrt{\frac{\mu_i}{\rho_i}}. \tag{10}$$

When wave speeds are supplied directly, they enter the same graph construction without the material conversion.

Let $\mathcal{C}_q^r = \{i \in V : \phi_r(i) = q\}$ be the fine nodes assigned to coarse node q. Coordinates and speeds are pooled before travel times are evaluated,

$$\boldsymbol{x}_q^r = \frac{1}{|\mathcal{C}_q^r|} \sum_{i \in \mathcal{C}_q^r} \boldsymbol{x}_i, \qquad c_q^{r,m} = \frac{1}{|\mathcal{C}_q^r|} \sum_{i \in \mathcal{C}_q^r} c_i^m. \tag{11}$$

Coarse edges connect distinct clusters linked by a fine edge, with duplicate edges removed. At each resolution, harmonic endpoint speeds and source-to-node arrivals are

$$d_{ij}^r = \|\boldsymbol{x}_i^r - \boldsymbol{x}_j^r\|_2, \qquad c_{ij}^{r,m} = \frac{2}{(c_i^{r,m})^{-1} + (c_j^{r,m})^{-1}}, \tag{12}$$

$$\tau_{ij}^{r,m} = \frac{d_{ij}^r}{c_{ij}^{r,m}}, \qquad a_j^{r,m} = \min_{\pi : S^r \rightsquigarrow j} \sum_{(u,v) \in \pi} \tau_{uv}^{r,m}. \tag{13}$$

Here $S^r = \{\phi_r(s) : s \in S\}$. Source arrivals are zero, and unreachable nodes have infinite arrival time. Recomputing these quantities on each graph makes coarse connections retain their own distance- and speed-dependent delays.

A.2 Gated Message Passing and Multiresolution Fusion

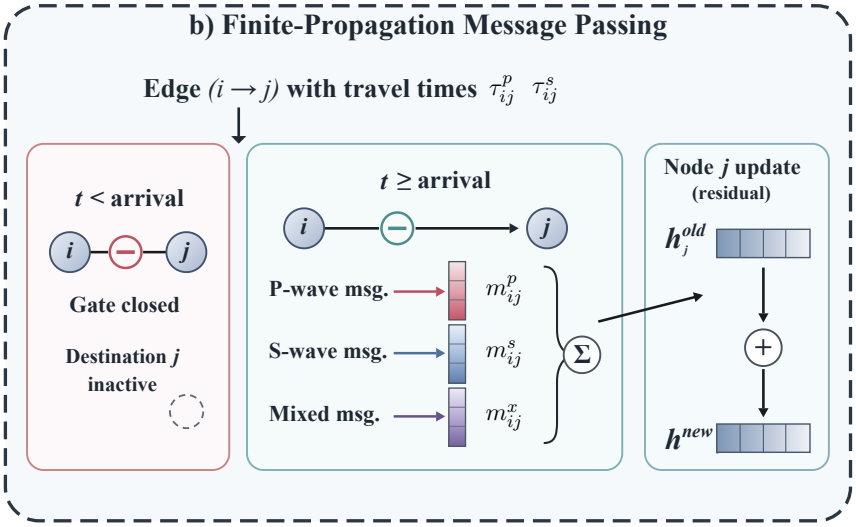

Figure 8: Finite-propagation message passing. Arrival gates weight separate P, S, and mixed message branches before aggregation and a residual node update. The closed/open cases illustrate hard gating, while soft gating makes the transition continuous around the arrival time.

Figure 8 details the propagation block. At resolution r and iteration k, the branch input and availability margin are

$$\boldsymbol{z}_{ij}^{r,k} = [\boldsymbol{h}_i^{r,k}, \boldsymbol{h}_j^{r,k}, \boldsymbol{e}_{ij}^r, \tau_{ij}^{r,p}, \tau_{ij}^{r,s}], \tag{14}$$

$$\Delta_{ij}^{r,m}(t) = t - a_i^{r,m} - \tau_{ij}^{r,m}. \tag{15}$$

Applying the gate in Eq. 5 to this margin gives $g_{ij}^{r,m}(t)$. Write $g_{ij}^{r,x} = \max(g_{ij}^{r,p}, g_{ij}^{r,s})$ for the mixed branch. For incoming neighbors $\mathcal{N}_j^r$, the branch messages, mean aggregation, and residual update are

$$\boldsymbol{m}_{ij}^{r,b,k} = g_{ij}^{r,b}(t)\, f_b^r(\boldsymbol{z}_{ij}^{r,k}), \qquad b \in \{p, s, x\}, \tag{16}$$

$$\overline{\boldsymbol{m}}_j^{r,b,k} = \frac{\sum_{i \in \mathcal{N}_j^r} \boldsymbol{m}_{ij}^{r,b,k}}{\max(1, |\mathcal{N}_j^r|)}, \tag{17}$$

$$\boldsymbol{h}_j^{r,k+1} = \mathrm{LayerNorm}\Big(\boldsymbol{h}_j^{r,k} + U_r([\boldsymbol{h}_j^{r,k}, \overline{\boldsymbol{m}}_j^{r,p,k}, \overline{\boldsymbol{m}}_j^{r,s,k}, \overline{\boldsymbol{m}}_j^{r,x,k}])\Big). \tag{18}$$

The functions f_b^r and U_r are MLPs. A closed hard gate removes the corresponding inter-node message. With $\eta > 0$, the sigmoid instead attenuates pre-arrival messages, as illustrated in Figure 3.

Each scale starts from the mean of the encoded fine-node states in its clusters. After K propagation iterations, states are gathered back to their fine nodes and fused,

$$\boldsymbol{h}_q^{r,0} = \frac{1}{|\mathcal{C}_q^r|} \sum_{i \in \mathcal{C}_q^r} \boldsymbol{h}_i^{(0)}, \tag{19}$$

$$\widehat{\boldsymbol{y}}_i = D\Big(F\Big([\boldsymbol{h}_{\phi_1(i)}^{1,K}, \ldots, \boldsymbol{h}_{\phi_R(i)}^{R,K}]\Big)\Big). \tag{20}$$

Here R is the number of resolutions, F is the fusion MLP, and D is the task decoder. This gather-and-fuse operation preserves a fine-node output while combining neighborhoods represented at different scales.

A.3 Causal Load Encoding Diagnostics

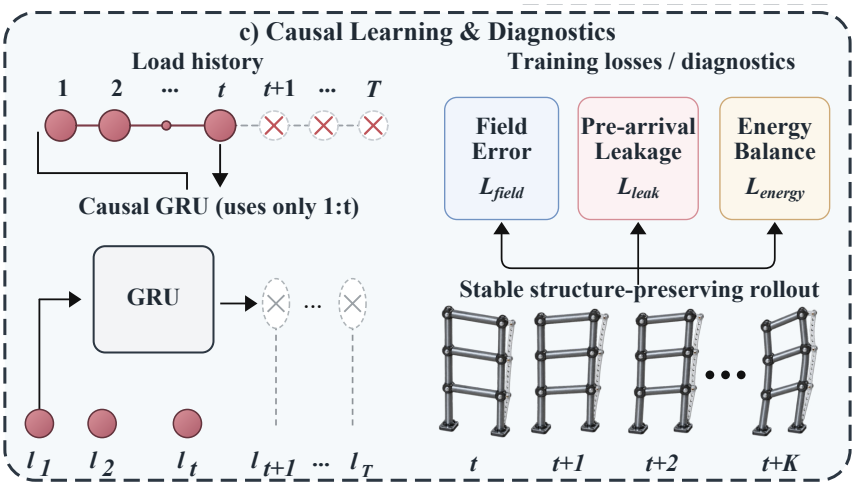

Figure 9: CLE diagnostics. The GRU processes only observed load samples. Field error measures predictive accuracy, pre-arrival leakage measures premature response, and the optional energy term evaluates balance when the required energy quantities are available. The rollout illustrates repeated response prediction.

Figure 9 separates the temporal input path from the quantities used to assess predictions. For a query at sample t, the encoder receives the prefix $\ell_i(1{:}t)$,

$$\boldsymbol{q}_{i,s} = \mathrm{GRU}(\ell_i(s), \boldsymbol{q}_{i,s-1}), \quad 1 \le s \le t, \qquad \boldsymbol{h}_i^{\ell} = \boldsymbol{q}_{i,t}. \tag{21}$$

The context enters the node encoder in Eq. (3). With all other inputs held constant, two load sequences sharing this prefix produce the same deterministic prediction,

$$\ell(1{:}t) = \widetilde{\ell}(1{:}t) \implies \widehat{\boldsymbol{y}}(t;\ell) = \widehat{\boldsymbol{y}}(t;\widetilde{\ell}). \tag{22}$$

This property underlies the future-only intervention in Table 3. Appendix A.4 gives the access proof.

For response arrays $\widehat{\boldsymbol{y}}$ and $\boldsymbol{y}$, the field error and leakage diagnostic are

$$\mathcal{L}_{\mathrm{field}} = \frac{\|\widehat{\boldsymbol{y}} - \boldsymbol{y}\|_2}{\max(\|\boldsymbol{y}\|_2, \varepsilon)}, \tag{23}$$

$$\mathcal{L}_{\mathrm{leak}} = \frac{\sum_{n,i} \mathbf{1}[t_n < a_i - \delta]\, \|\widehat{\boldsymbol{y}}_i(t_n)\|_2^2}{\max(\sum_{n,i} \|\widehat{\boldsymbol{y}}_i(t_n)\|_2^2, \varepsilon)}. \tag{24}$$

Here $\varepsilon > 0$ stabilizes normalization, a_i is the reference arrival in the controlled system, and $\delta \ge 0$ is a pre-arrival margin. The leakage ratio measures response energy before arrival, whereas the future intervention measures sensitivity to unavailable load samples.

When kinetic energy T, strain energy V, accumulated external work W, and accumulated dissipation Q are available, the optional energy diagnostic is

$$\Delta\mathcal{E}_n = T_n + V_n - T_0 - V_0, \qquad B_n = W_n - W_0 - Q_n, \tag{25}$$

$$\mathbf{e}_{\mathcal{E}} = (\Delta\mathcal{E}_n)_n, \qquad \mathbf{b} = (B_n)_n, \tag{26}$$

$$\mathcal{L}_{\mathrm{energy}} = \frac{\|\mathbf{e}_{\mathcal{E}} - \mathbf{b}\|_2}{\max(\|\mathbf{b}\|_2, \varepsilon)}. \tag{27}$$

The energy diagnostic requires these additional physical quantities. The reported mechanism study uses pre-arrival energy, arrival MAE, and future-only interventions to evaluate propagation and history access separately.

A.4 FINITE-PROPAGATION CAUSALITY PROOF

We write the proof as an access calculation at query time t.

Notation. For scale r, let $\mathcal{G}^r = (V^r, E^r)$ and $S^r \subseteq V^r$. For $m \in \{p, s\}$,

$$a_i^{r,m} = \min_{\pi: S^r \rightsquigarrow i} \sum_{(u,v)\in\pi} \tau_{uv}^{r,m}, \qquad \tau_{uv}^{r,m} \geq 0, \tag{28}$$

$$\Delta_{ij}^{r,m}(t) = t - a_i^{r,m} - \tau_{ij}^{r,m}, \tag{29}$$

$$g_{ij}^{r,m}(t) = \begin{cases} \mathbf{1}[\Delta_{ij}^{r,m}(t) \geq 0], & \eta \leq 0, \\ \sigma(\Delta_{ij}^{r,m}(t)/\eta), & \eta > 0, \end{cases} \tag{30}$$

$$\mathcal{R}^{r,m}(t) = \{i \in V^r : a_i^{r,m} \leq t\}. \tag{31}$$

If no path exists, $a_i^{r,m} = \infty$. For the mixed branch $b = x$,

$$\Delta_{ij}^{r,x}(t) = \max\{\Delta_{ij}^{r,p}(t), \Delta_{ij}^{r,s}(t)\}, \tag{32}$$

$$g_{ij}^{r,x}(t) = \max\{g_{ij}^{r,p}(t), g_{ij}^{r,s}(t)\}, \tag{33}$$

$$\mathcal{R}^{r,*}(t) = \mathcal{R}^{r,p}(t) \cup \mathcal{R}^{r,s}(t). \tag{34}$$

For $b \in \{p, s, x\}$, write

$$\boldsymbol{m}_{ij}^{r,k,b}(t) = g_{ij}^{r,b}(t)\, f_b\Big([\boldsymbol{h}_i^{r,k}(t), \boldsymbol{h}_j^{r,k}(t), \boldsymbol{e}_{ij}^r, \tau_{ij}^{r,p}, \tau_{ij}^{r,s}]\Big). \tag{35}$$

Lemma 1 (Future-load invariance)**.** *Under Assumption 1, if*

$$X \equiv_t X' \quad \Longleftrightarrow \quad \begin{cases} \mathcal{G}_X^r = \mathcal{G}_{X'}^r, \quad S_X^r = S_{X'}^r, \quad \tau_{X,ij}^{r,m} = \tau_{X',ij}^{r,m}, \\ \ell_i^X(q) = \ell_i^{X'}(q), \quad \forall i,\ q \leq t, \end{cases} \tag{36}$$

while $\ell_i^X(q)$ and $\ell_i^{X'}(q)$ may differ for $q > t$, then

$$\boldsymbol{h}_{X,j}^{r,k}(t) = \boldsymbol{h}_{X',j}^{r,k}(t) \quad \forall r, k, j, \qquad \widehat{\boldsymbol{y}}_X(t) = \widehat{\boldsymbol{y}}_{X'}(t). \tag{37}$$

Proof.

$$\ell_i^X(1{:}t) = \ell_i^{X'}(1{:}t) \Rightarrow \boldsymbol{h}_i^{\ell,X}(t) = \boldsymbol{h}_i^{\ell,X'}(t), \tag{38}$$

$$\mathcal{G}_X^r = \mathcal{G}_{X'}^r,\ \tau_X^{r,m} = \tau_{X'}^{r,m},\ S_X^r = S_{X'}^r \Rightarrow a_{X,i}^{r,m} = a_{X',i}^{r,m},\ g_{X,ij}^{r,b}(t) = g_{X',ij}^{r,b}(t). \tag{39}$$

The encoder therefore gives

$$\boldsymbol{h}_{X,j}^{r,0}(t) = \boldsymbol{h}_{X',j}^{r,0}(t) \qquad \forall r, j. \tag{40}$$

Assume $\boldsymbol{h}_{X,j}^{r,k}(t) = \boldsymbol{h}_{X',j}^{r,k}(t)$ for all j. Then by Eq. equation 35,

$$\boldsymbol{m}_{X,ij}^{r,k,b}(t) = \boldsymbol{m}_{X',ij}^{r,k,b}(t) \qquad \forall (i,j) \in E^r,\ b \in \{p, s, x\}. \tag{41}$$

For the deterministic aggregation and update maps A_j^r, U^r,

$$\bar{\boldsymbol{m}}_{X,j}^{r,k}(t) = A_j^r\Big(\{\boldsymbol{m}_{X,ij}^{r,k,b}(t)\}_{i,b}\Big) = A_j^r\Big(\{\boldsymbol{m}_{X',ij}^{r,k,b}(t)\}_{i,b}\Big) = \bar{\boldsymbol{m}}_{X',j}^{r,k}(t), \tag{42}$$

$$\boldsymbol{h}_{X,j}^{r,k+1}(t) = U^r\Big(\boldsymbol{h}_{X,j}^{r,k}(t), \bar{\boldsymbol{m}}_{X,j}^{r,k}(t)\Big) = U^r\Big(\boldsymbol{h}_{X',j}^{r,k}(t), \bar{\boldsymbol{m}}_{X',j}^{r,k}(t)\Big) = \boldsymbol{h}_{X',j}^{r,k+1}(t). \tag{43}$$

Local restriction/prolongation, fusion F, and decoder D preserve equality.

$$\{\boldsymbol{h}_X^{r,K_r}\}_r = \{\boldsymbol{h}_{X'}^{r,K_r}\}_r \Rightarrow \boldsymbol{z}_X(t) = F(\{\boldsymbol{h}_X^{r,K_r}\}_r) = \boldsymbol{z}_{X'}(t) \Rightarrow \widehat{\boldsymbol{y}}_X(t) = D(\boldsymbol{z}_X(t)) = D(\boldsymbol{z}_{X'}(t)). \tag{44}$$

□

Proposition 1 (Hard-gated finite-propagation causality)**.** *Assume $\eta \leq 0$, so that $g_{ij}^{r,m}(t) = \mathbf{1}[\Delta_{ij}^{r,m}(t) \geq 0]$ for $m \in \{p, s\}$. Then the following statements hold at every graph scale r and query time t.*

1. *A nonzero p- or s-branch edge message on $i \to j$ is possible only if*
$$a_i^{r,m} + \tau_{ij}^{r,m} \le t. \tag{45}$$
A nonzero mixed-branch edge message is possible only if this inequality holds for at least one of $m = p$ or $m = s$.

2. *Let a source-load intervention change only source load histories $\{\ell_s(q) : s \in S^r,\ q \le t\}$ while keeping geometry, materials, graph structure, source set, and non-source loads unchanged. If $j \notin \mathcal{R}^{r,*}(t)$, then the state of graph element j after any finite number of intra-scale message-passing layers is invariant to that source-load intervention.*

Proof.

$$g_{ij}^{r,m}(t) = 1 \iff \Delta_{ij}^{r,m}(t) = t - a_i^{r,m} - \tau_{ij}^{r,m} \ge 0 \qquad \iff a_i^{r,m} + \tau_{ij}^{r,m} \le t, \tag{46}$$

$$\Delta_{ij}^{r,m}(t) < 0 \Longrightarrow g_{ij}^{r,m}(t) = 0 \Longrightarrow \boldsymbol{m}_{ij}^{r,k,m}(t) = \mathbf{0}. \tag{47}$$

For $b = x$,

$$\boldsymbol{m}_{ij}^{r,k,x}(t) \neq \mathbf{0} \Longrightarrow \max\{g_{ij}^{r,p}(t), g_{ij}^{r,s}(t)\} = 1 \Longrightarrow \exists m \in \{p,s\} : a_i^{r,m} + \tau_{ij}^{r,m} \le t. \tag{48}$$

Let $\delta \boldsymbol{h}_j^{r,k}(t)$ denote the change induced by the source-load intervention. The target is

$$j \notin \mathcal{R}^{r,*}(t) \Longrightarrow \delta \boldsymbol{h}_j^{r,k}(t) = \mathbf{0} \qquad \forall k < \infty. \tag{49}$$

At $k = 0$,

$$\delta \boldsymbol{h}_j^{r,0}(t) = \mathbf{0} \quad \text{for} \quad j \notin \mathcal{R}^{r,*}(t), \tag{50}$$

because the intervention is supported only on S^r and

$$s \in S^r \Rightarrow a_s^{r,p} = a_s^{r,s} = 0 \Rightarrow s \in \mathcal{R}^{r,*}(t). \tag{51}$$

Assume Eq. equation 49 holds at layer k. For $j \notin \mathcal{R}^{r,*}(t)$ and an incoming edge $i \to j$,

$$\begin{aligned} i \in \mathcal{R}^{r,m}(t),\ a_i^{r,m} + \tau_{ij}^{r,m} \le t &\Rightarrow a_j^{r,m} \le a_i^{r,m} + \tau_{ij}^{r,m} \le t \\ &\Rightarrow j \in \mathcal{R}^{r,m}(t) \subseteq \mathcal{R}^{r,*}(t), \end{aligned} \tag{52}$$

which is impossible. Hence

$$j \notin \mathcal{R}^{r,*}(t),\ i \in \mathcal{R}^{r,*}(t) \Longrightarrow g_{ij}^{r,p}(t) = g_{ij}^{r,s}(t) = g_{ij}^{r,x}(t) = 0. \tag{53}$$

Therefore every incoming perturbation term satisfies

$$\delta \boldsymbol{m}_{ij}^{r,k,b}(t) = \mathbf{0}, \qquad b \in \{p,s,x\}, \tag{54}$$

either because $\delta \boldsymbol{h}_i^{r,k}(t) = \mathbf{0}$ by induction or because Eq. equation 53 sets the gate to zero. Thus

$$\delta \bar{\boldsymbol{m}}_j^{r,k}(t) = \mathbf{0}, \tag{55}$$

$$\delta \boldsymbol{h}_j^{r,k+1}(t) = U^r\left(\boldsymbol{h}_j^{r,k} + \delta \boldsymbol{h}_j^{r,k}, \bar{\boldsymbol{m}}_j^{r,k} + \delta \bar{\boldsymbol{m}}_j^{r,k}\right) - U^r\left(\boldsymbol{h}_j^{r,k}, \bar{\boldsymbol{m}}_j^{r,k}\right) = \mathbf{0}. \tag{56}$$

Induction proves Eq. equation 49. □

Proposition 2 (Soft-gate pre-arrival leakage bound)**.** *Assume $\eta > 0$. Suppose that, on the bounded feature domain reached by the trained model, each branch function satisfies $\|f_b(\cdot)\| \le B_b$ for $b \in \{p,s,x\}$, and the update, scale-transfer, fusion, and decoder maps have finite Lipschitz constants. For a branch-edge term whose soft-gate margin obeys $\Delta_{ij}^{r,b}(t) \le -\gamma$ with $\gamma > 0$, its pre-arrival message magnitude is bounded by*

$$\|\boldsymbol{m}_{ij}^{r,b}(t)\| \le B_b \exp(-\gamma/\eta). \tag{57}$$

Consequently, over a finite multiresolution graph and a finite number of layers, the total output contribution caused by all branch-edge terms with margin at most $-\gamma$ is bounded by

$$\|\widehat{\boldsymbol{y}}_{\text{leak}}(t)\| \le C_{\mathcal{G}} B_{\max} \exp(-\gamma/\eta), \tag{58}$$

where $B_{\max} = \max_b B_b$ and $C_{\mathcal{G}} < \infty$ depends on the number of scales, edges, layers, aggregation norms, and Lipschitz constants, but not on η or γ. Without a uniform margin, the sharper finite-sum bound is obtained by replacing $\exp(-\gamma/\eta)$ with the corresponding gate values $\sigma(\Delta_{ij}^{r,b}(t)/\eta)$ and summing over the closed branch-edge terms.

Proof. For any scalar $z \le -\gamma/\eta < 0$,

$$\sigma(z) = \frac{1}{1+\exp(-z)} \le \exp(z) \le \exp(-\gamma/\eta). \tag{59}$$

For $b = p$ or $b = s$, let $z = \Delta_{ij}^{r,b}(t)/\eta$.

$$\begin{aligned} \|\boldsymbol{m}_{ij}^{r,b}(t)\| &= \sigma(\Delta_{ij}^{r,b}(t)/\eta)\, \|f_b(\cdot)\| \\ &\le B_b \exp(-\gamma/\eta). \end{aligned} \tag{60}$$

For $b = x$,

$$\begin{aligned} g_{ij}^{r,x}(t) &= \max\{\sigma(\Delta_{ij}^{r,p}/\eta), \sigma(\Delta_{ij}^{r,s}/\eta)\} \\ &= \sigma\big(\max\{\Delta_{ij}^{r,p}, \Delta_{ij}^{r,s}\}/\eta\big) = \sigma(\Delta_{ij}^{r,x}/\eta), \end{aligned} \tag{61}$$

so Eq. equation 60 also holds when $\Delta_{ij}^{r,x}(t) \le -\gamma$.

Let

$$\epsilon_\gamma = B_{\max} \exp(-\gamma/\eta), \qquad \mathcal{E}_\gamma^{r,k} = \{(i,j,b) : \Delta_{ij}^{r,b}(t) \le -\gamma\}. \tag{62}$$

If aggregation at scale r has operator norm at most α^r and the update map has Lipschitz constant L_U^r, the state perturbation caused by closed branch-edge terms satisfies

$$\delta^{r,k+1} \le L_U^r \delta^{r,k} + L_U^r \alpha^r |\mathcal{E}_\gamma^{r,k}| \epsilon_\gamma. \tag{63}$$

Unrolling K_r layers gives

$$\delta^{r,K_r} \le \left(\sum_{k=0}^{K_r-1} (L_U^r)^{K_r-k} \alpha^r |\mathcal{E}_\gamma^{r,k}| \right) \epsilon_\gamma = C_r \epsilon_\gamma. \tag{64}$$

With finite scale-transfer constants L_T^r, fusion constant L_F, and decoder constant L_D,

$$\begin{aligned} \|\widehat{\boldsymbol{y}}_{\text{leak}}(t)\| &\le L_D L_F \sum_r \left(\prod_{q \in \mathcal{P}(r)} L_T^q \right) \delta^{r,K_r} \\ &\le \underbrace{L_D L_F \sum_r \left(\prod_{q \in \mathcal{P}(r)} L_T^q \right) C_r}_{C_{\mathcal{G}}}\, B_{\max} \exp(-\gamma/\eta). \end{aligned} \tag{65}$$

If the margins differ, replace ϵ_γ in Eqs. equation 63–equation 64 by

$$\epsilon_{ijb}^{r,k} = B_b\, \sigma(\Delta_{ij}^{r,b}(t)/\eta), \tag{66}$$

and keep the resulting finite sum. □

Corollary 1 (No zero-time multiresolution shortcut)**.** *Under the hard-gate setting, every nonzero coarse or fine graph message in FPMR-NO must satisfy the arrival inequality computed at its own graph scale. Under the soft-gate setting, any message sent before that arrival condition has the exponential leakage bound in Proposition 2. Therefore multiresolution coarsening can shorten the number of computational message-passing steps, but it does not turn a long-range coarse edge into a zero-time physical communication path.*

Proof.

$$\forall r,\ (i,j) \in E^r,\ m \in \{p,s\} : \quad \boldsymbol{m}_{ij}^{r,k,m}(t) \ne \mathbf{0} \Rightarrow a_i^{r,m} + \tau_{ij}^{r,m} \le t \qquad (\eta \le 0), \tag{67}$$

$$\Delta_{ij}^{r,b}(t) \le -\gamma \Rightarrow \|\boldsymbol{m}_{ij}^{r,k,b}(t)\| \le B_b e^{-\gamma/\eta} \qquad (\eta > 0). \tag{68}$$

The same formulas are evaluated after each coarsening step.

$$\tau_{ij}^{r,m} = \frac{\|\boldsymbol{x}_i^r - \boldsymbol{x}_j^r\|}{2} \left(\frac{1}{c_i^{r,m}} + \frac{1}{c_j^{r,m}} \right), \qquad a_j^{r,m} = \min_{\pi : S^r \rightsquigarrow j} \sum_{(u,v) \in \pi} \tau_{uv}^{r,m}. \tag{69}$$

Thus a coarse edge changes the graph distance, but the gate still contains $a_i^{r,m} + \tau_{ij}^{r,m} \le t$; before this inequality holds, the hard message is zero and the soft message is exponentially bounded. □

B EXPERIMENTAL DETAILS

B.1 DATASETS

We evaluate FPMR-NO on two structural-dynamics datasets with complementary observation settings. DeformingPlate follows the official AMG one-step stress setting on variable-size tetrahedral meshes and is used to test irregular-geometry stress prediction. ETH contains laboratory measurements from a four-story shear-type frame and is used to test sparse-sensor impact-response prediction. The formal dataset scales and prediction targets are summarized in Table 5; the detailed ETH windowing and preprocessing procedure is given in the subsequent data-processing subsection.

Table 5: Summary of the datasets used in the formal experiments. The split sizes are counts of trajectories for DeformingPlate and counts of supervised impact events for ETH.

Dataset	Source and structure	Scale and split	Model input	Prediction target
DeformingPlate	Variable-size tetrahedral meshes under the official AMG stress setting (Li et al., 2025)	1,200 train, 100 validation, and 100 test trajectories	Mesh geometry and current nodal state	One-step nodal stress residual
ETH	Laboratory four-story shear frame with impact and vibration measurements (Brambilla et al., 2025)	32 records and 256 events; 192 train, 32 validation, and 32 test events	Length-4096 hammer history, six acceleration histories, and sensor/impact metadata	Six sparse accelerometer response histories

B.2 COMPARATIVE METHODS

LSTM (Hochreiter & Schmidhuber, 1997). The LSTM baseline processes the hammer or response sequence recurrently with gated memory cells. Its input, forget, and output gates provide a standard temporal model for comparing sequence-only dynamics modeling.

Transformer (Vaswani et al., 2017). The Transformer baseline uses self-attention and feed-forward blocks to model dependencies across the input sequence without recurrent state updates. It provides a general sequence-attention reference for the ETH response task.

MeshGraphNet (Pfaff et al., 2021). MeshGraphNet is an encode–process–decode graph simulator. It represents mesh elements and nodes as graph features, propagates local messages along mesh edges, and updates the node states through repeated graph-network blocks.

DeepONet (Lu et al., 2021). DeepONet separates operator learning into a branch network that encodes sampled input functions and a trunk network that encodes query locations. Their combined representation produces the output function at the requested sensor or spatial coordinates.

GNOT (Hao et al., 2023). GNOT is a transformer-based neural operator for irregular meshes and multiple input functions. Its heterogeneous normalized attention handles heterogeneous inputs, while geometric gating provides a soft decomposition for multiscale interactions.

Transolver (Wu et al., 2024). Transolver groups mesh points into learnable slices according to latent physical states and applies attention to physics-aware slice tokens. This converts point-level interaction into a geometry-general physical-state representation with efficient attention.

Poseidon (Herde et al., 2024). Poseidon is a pretrained PDE foundation model based on a multiscale operator Transformer with time-conditioned normalization. It uses large-scale pretraining and semigroup-based data construction to transfer operator representations to downstream PDE tasks.

AMG (Li et al., 2025). AMG constructs a multi-graph neural operator for arbitrary geometries. It combines scale graphs for multiresolution information with a physics graph and uses graph-based attention to model dependencies across irregular domains.

RIGNO (Mousavi et al., 2025). RIGNO is a graph neural operator that maps between point-cloud inputs and outputs through a downsampled regional mesh. Its multiscale regional interactions target resolution-invariant operator learning on arbitrary domains.

Transolver++ (Luo et al., 2025). Transolver++ extends slice-based physical-state modeling to million-scale geometries. It adds highly parallel computation and local adaptive mechanisms so that the solver can scale with the number of mesh points.

RNO (Liu et al., 2026). RNO introduces spectral derivative structure into operator layers to combine local and global information. It serves as a spectral neural-operator comparison for the ETH response task.

B.3 Source-Load Intervention Setup

Figure 5 uses a separate single-scale controlled experiment, distinct from the three-seed aggregate in Table 3. A one-dimensional step wave propagates from position zero with true speed $c = 1$ on a complete directed graph of 16 uniformly spaced nodes in $[0, 1]$. The FPMR-NO and w/o PG models each contain 31,105 parameters, hidden width 32, and one message-passing step. Both use seed 2024, 4096 training samples, 1024 validation samples, batch size 128, and 100 training epochs. AdamW uses learning rate 0.002, weight decay 10^{-5}, and cosine decay; the model state with the best validation loss is used, with early-stopping patience 20. This visualization run uses PyTorch 2.1.1 with CUDA 12.1 on an NVIDIA GeForce RTX 3050, separately from the A40 measurements reported for the formal benchmarks.

The case is specified before training. Predictions are evaluated at 241 uniformly spaced times in $[0, 1.45]$ while source amplitude changes from 1.0 to 1.2, with all other inputs held constant. The source-load influence and its pre-arrival RMS are

$$I_j(t) = \frac{|\hat{y}_j(t; 1.2) - \hat{y}_j(t; 1.0)|}{0.2}, \tag{70}$$

$$\mathcal{P} = \{(j, t) : t < x_j/c\}, \qquad R_{\mathrm{pre}} = \left(\frac{1}{|\mathcal{P}|} \sum_{(j,t)\in\mathcal{P}} I_j(t)^2 \right)^{1/2}. \tag{71}$$

The gated model uses softness $\eta = 0.0125$. At inference, its assumed gate speed is set to 0.5, 1, or 2 without retraining, while the target wave and all non-speed inputs remain unchanged. The ungated counterpart is trained independently with the gate disabled. The display floor 10^{-5} is used only for logarithmic plotting, not metric computation.

Correct speed concentrates pre-arrival sensitivity in a narrow band near the arrival boundary, whereas ungated sensitivity is more spatially dispersed. In this case, R_{pre} is nevertheless 6.69×10^{-2} with gating versus 4.01×10^{-2} without it because the localized band contributes substantially to the squared sum. Response RMS, which measures the predicted response rather than its sensitivity to amplitude, is 1.43×10^{-2} versus 3.94×10^{-2}. These are different diagnostics. This figure shows the spatial organization and speed dependence of learned influence, not a uniform reduction of integrated influence or a multiscale test.

B.4 Additional Experimental Figures

Figure 10 shows a predetermined DeformingPlate test trajectory.

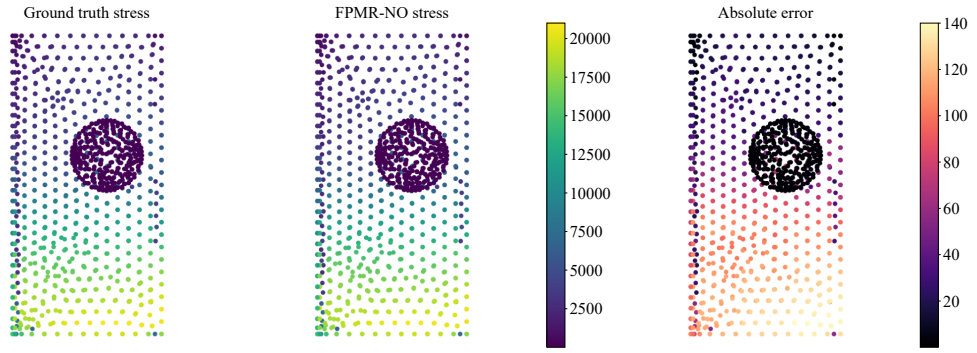

Figure 10: Representative DeformingPlate stress prediction on the original irregular mesh. The predetermined trajectory compares nodal stress and absolute error and was chosen before test evaluation rather than selected by error.

Figure 11 visualizes the controlled results in Table 3.

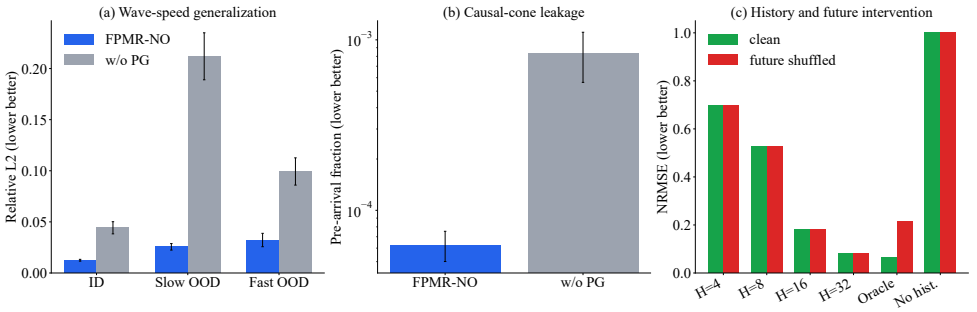

Figure 11: Controlled finite-propagation and causal-encoding evidence. The panels summarize travel-time gating, causal history length, and future-intervention behavior under identifiable synthetic conditions.

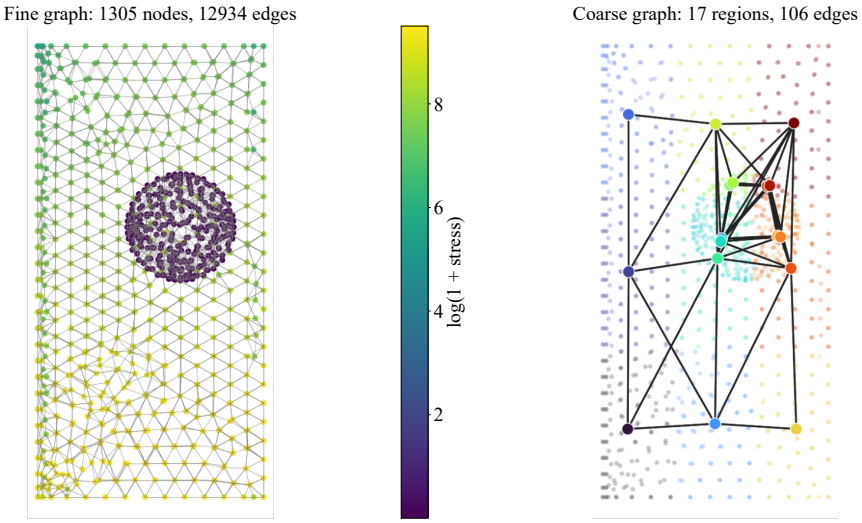

Figure 12: DeformingPlate multiresolution graph diagnostic. Fine tetrahedral adjacency and deterministic coarse regions are determined from the input geometry and do not use labels.

Figures 6, 13, 14, 15, 16, and 17 collect supplementary visual evidence for accuracy-efficiency trade-offs, qualitative stress contours, frequency response, robustness, temporal generalization, and scaling.

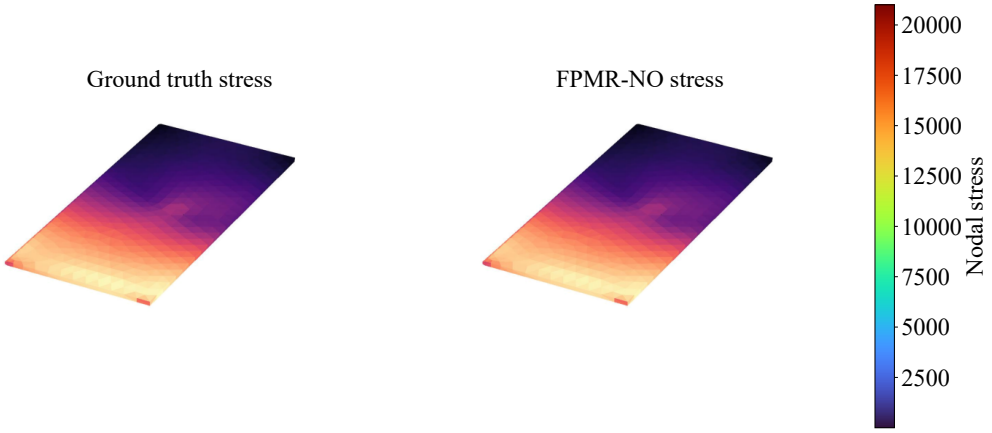

Figure 13: Computer-aided engineering (CAE)-style Plate stress contours for the same test case. Ground truth and FPMR-NO prediction use the same nodal-stress color scale and camera, enabling direct comparison of the high-stress pattern.

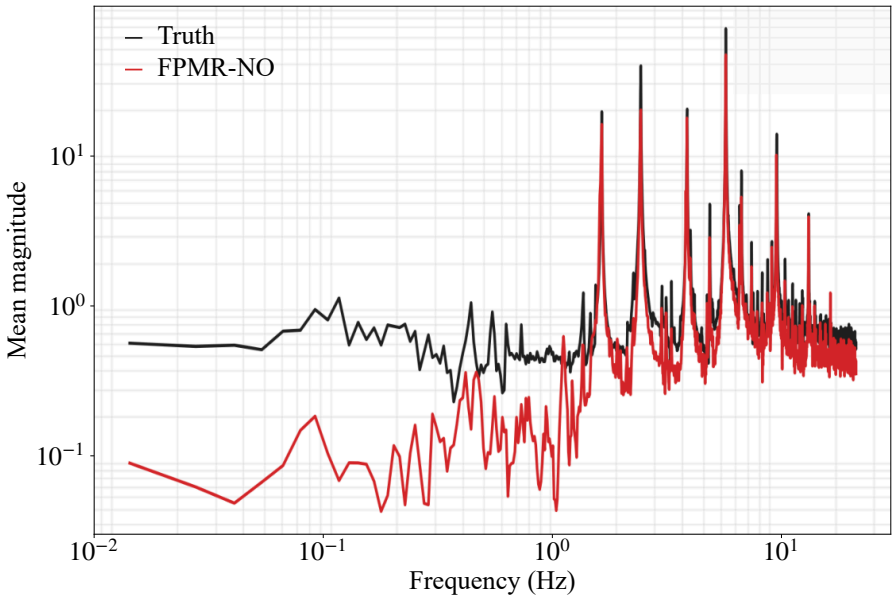

Figure 14: Channel-averaged ETH spectra for the same test event, showing that the time-domain response comparison is accompanied by frequency-response evidence.

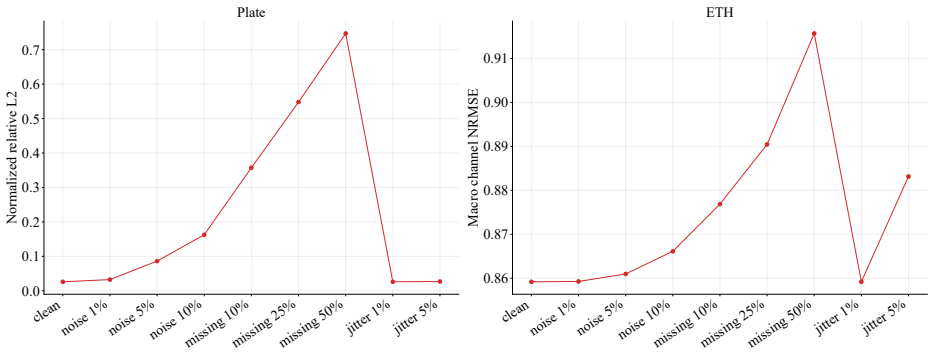

Figure 15: Robustness grid over noise, missingness, and geometry perturbations. The grid highlights that sparse ETH response transfer and dense Plate stress prediction have different sensitivities.

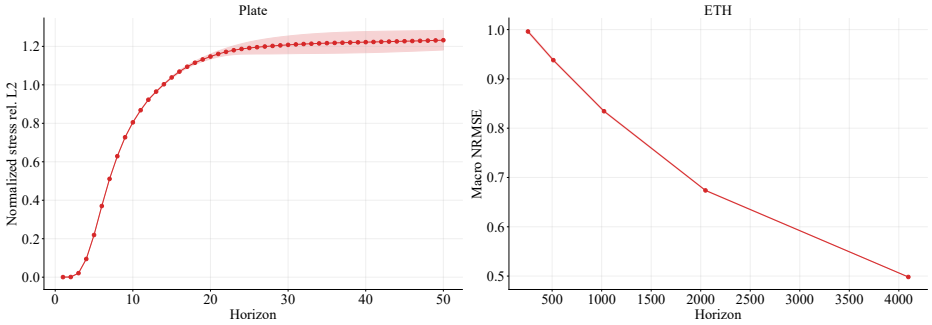

Figure 16: Validation-selected model-state temporal diagnostics. Plate uses autoregressive rollout; ETH measures channel-averaged NRMSE on training-standardized response prefixes, rather than recursive extrapolation. Both panels show unscaled errors.

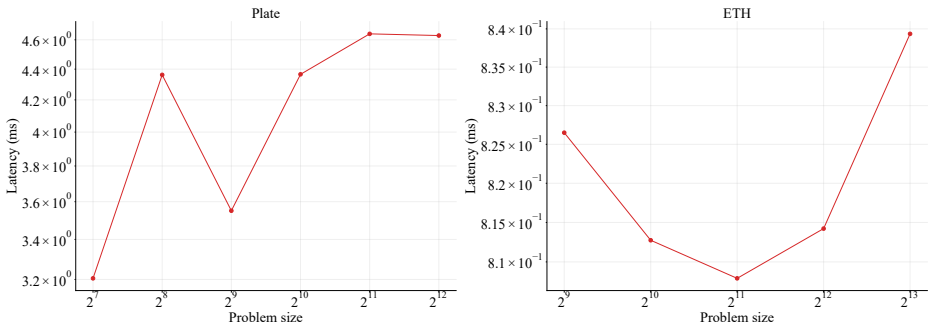

Figure 17: Synchronized FPMR-NO inference scaling after warm-up. The measurements separate batch-size-one latency from model-selection metrics.

B.5 ETH Data-Processing Details

The ETH impact task is built from 32 physical records with eight detected hammer impacts per record, yielding 256 supervised events. Each event is windowed with 82 pre-trigger and 8190 post-trigger raw samples, then represented as a length-4096 hammer history and six acceleration histories. Splits are group-disjoint by roving sensor-node pair and stratified by impact direction. The split contains 192 training events, 32 validation events, and 32 testing events. The held-out test groups are not used for event normalization, cross-spectral initialization, hyperparameter selection, or early stopping.

Normalization is fitted on training events only. The hammer input is standardized with the training input mean and standard deviation, and each acceleration channel is standardized with its own training-channel statistics; metrics are computed after inverse transformation to physical response units. Structural conditioning uses five guide-derived topology features per response channel. These features include normalized corner coordinates, normalized layer height, encoded sensor direction, and encoded impact direction. ETH TI estimates the hammer-to-accelerometer cross-spectral transfer on training data only, applies the selected spectral regularization, and truncates the singular value decomposition (SVD) transfer bank to the validation-selected 32 modes.

B.6 Parameter Sensitivity

We evaluate parameter sensitivity with one-factor-at-a-time sweeps on the validation split using an independent seed, 991. The test split is not accessed in this analysis. The selected default is the main validation-selected configuration, and each alternative changes only the named setting. Figure 18 reports the percentage change in the primary validation error relative to that default.

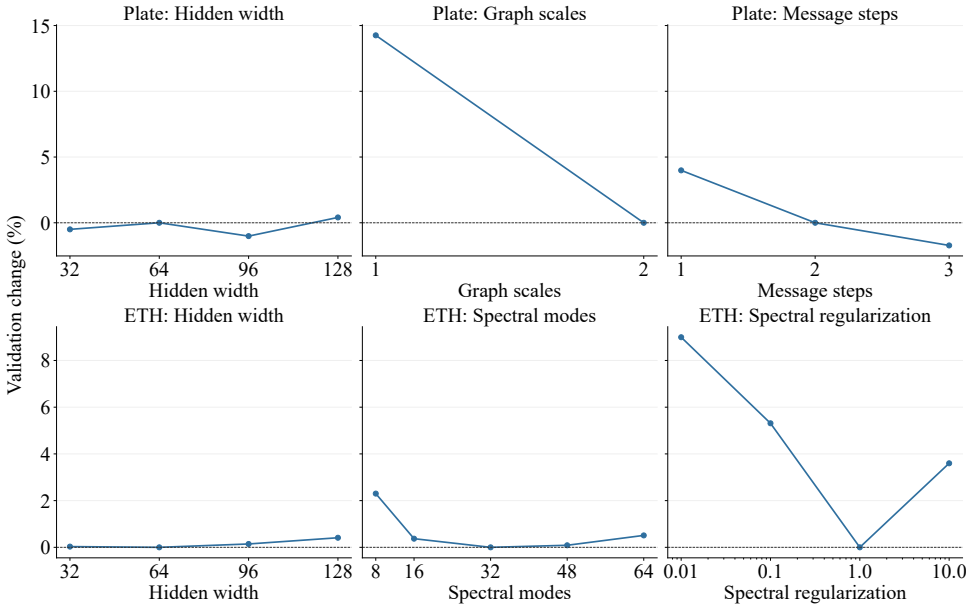

Figure 18: Validation-only parameter sensitivity curves. Each curve changes one model setting while keeping the remaining configuration unchanged. Values show percentage change in the primary validation error relative to the selected default, with positive values indicating higher error.

The Plate sweep is most sensitive to the number of graph scales. Removing the second scale increases validation error by 14.6%, whereas changing the hidden width or message depth produces smaller changes. On ETH, the spectral regularization coefficient has the clearest effect, with the lowest tested value increasing macro NRMSE by 9.0% relative to the selected default. The mode-count and hidden-width sweeps remain within a narrower range. These results indicate that the selected configuration is supported by a stable validation region, while multiresolution aggregation on Plate and spectral regularization on ETH are the most consequential settings.

B.7 Statistical Comparisons on ETH

We compare the three matched ETH seeds with a two-sided paired t-test. Figure 19 displays the comparisons that remain significant after Holm correction. The primary errors use the section-wide 10^{-2} coefficient scale.

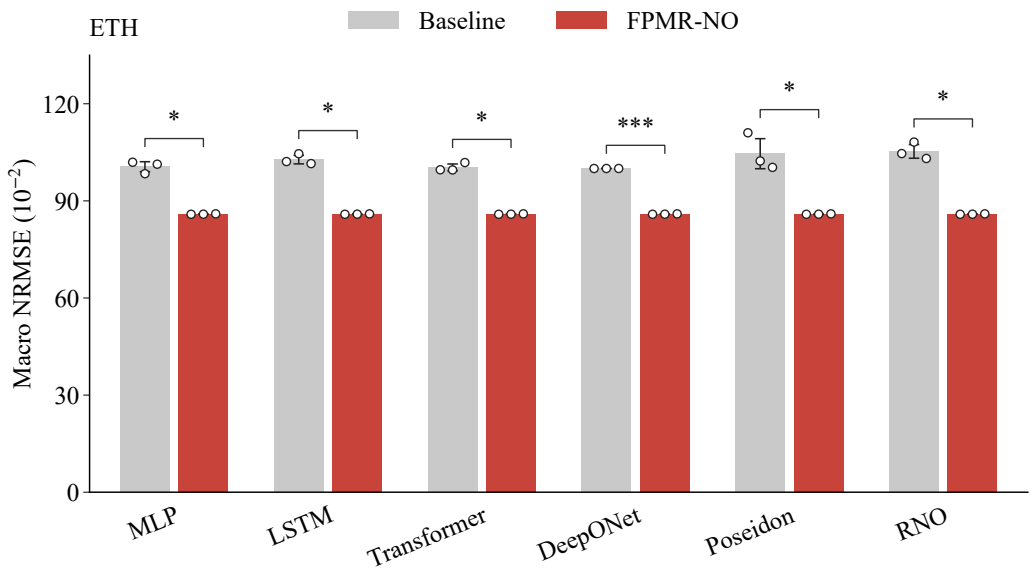

Figure 19: Significant paired t-test comparisons on ETH across three seeds. Gray and red bars show baseline and FPMR-NO means, error bars show population standard deviations, and hollow circles show individual runs. Each pair reuses the same three FPMR-NO runs. All six ETH comparisons have Holm-adjusted $p < 0.05$. Brackets show the adjusted levels, where $*$, $**$, and $***$ indicate $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively.

After Holm correction, FPMR-NO is significantly lower than all six listed ETH baselines, with adjusted p-values ranging from 0.0122 to 0.0301. The paired t-test supports the consistent ETH margin across the compared methods.

B.8 Data Efficiency and Extended Completion

Table 6: Three-seed low-data efficiency. Error metrics use the section-wide reporting scale, and standard deviations are shown to four decimals. Bold indicates the lower mean for each dataset and fraction.

Dataset	Comparator	Train frac.	Comparator error	FPMR-NO error	Rel. gain
Plate	Transolver	10%	2.98 ± 0.1108	$\mathbf{2.83 \pm 0.0431}$	4.93%
Plate	Transolver	25%	2.76 ± 0.0275	$\mathbf{2.72 \pm 0.0588}$	1.39%
Plate	Transolver	50%	2.69 ± 0.0270	$\mathbf{2.66 \pm 0.0847}$	0.81%
ETH	DeepONet	10%	100.06 ± 0.0384	$\mathbf{89.47 \pm 5.9478}$	10.59%
ETH	DeepONet	25%	100.02 ± 0.0118	$\mathbf{89.56 \pm 0.1118}$	10.46%
ETH	DeepONet	50%	100.01 ± 0.0082	$\mathbf{87.70 \pm 1.2117}$	12.31%

The updated experiment bundle completes all planned extended runs (95/95). The newly completed Plate low-data Transolver runs make the FPMR/Transolver comparison complete at 10%, 25%, and 50% training fractions. Table 6 reports the three-seed mean and standard deviation. FPMR-NO remains lower at every low-data fraction on Plate and keeps a larger margin over DeepONet in sparse ETH response transfer.