SONAR: SENSOR-CONDITIONED NEURAL OPERATORS WITH ADAPTIVE SPECTRAL ROTATION FOR FULL-FIELD PREDICTION ON MESHES

Anonymous authors
Paper under double-blind review

ABSTRACT

Existing physical-field prediction paradigms include autoregressive simulators and in-situ sensors reconstruction, which rely on complete prior field data or dense in-domain sensors and thus cannot be deployed under practical equipment operating conditions. Real engineering systems only provide sparse low-dimensional time-series observations. These observations mismatch target physical fields in both units and spatial locations, making it hard to infer high-dimensional field evolution from limited sensor measurements. This raises a critical challenge for high-precision physical field prediction of edge-side equipment: revealing how sparse heterogeneous observations propagate in high-dimensional field spaces and regulate global field distributions. Therefore, we define a more rigorous edge-side physics-AI task: causal next-step full-field prediction on unstructured meshes from short windows of heterogeneous observed state sequences. Duhamel–Volterra and observability analysis show that the predictable component of the field is fully spanned by causal convolutions of the observed states, unobservable field energy sets an error floor, and the optimal predictor admits a conditional spectral decomposition. We instantiate this framework in SONAR: observation history drives the rotation of mesh spectral basis fields generated by deep message passing and architecturally Gram-whitened; sensor-derived coefficients synthesize the global field, and a curvature-guided active-set residual repairs localized high-frequency errors. Experiments on elastic-beam, heat-plate, and mechanical-leg benchmarks with ablation, uncertainty, and robustness analyses show that SONAR achieves the best accuracy among compared methods under a matched information budget.

1 INTRODUCTION

High-fidelity simulation of physical fields such as stress in loaded structures, heat flux in thermal systems, and velocity in fluid domains is the computational backbone of engineering design, structural health monitoring, and closed-loop control. The per-step cost of classical numerical solvers remains prohibitive for real-time applications. Learned surrogates have emerged as a compelling alternative. Neural operators (Li et al., 2021; Kovachki et al., 2023) learn mappings between function spaces and generalize across problem instances, while learned simulators (Lam et al., 2023; Bi et al., 2023) roll out complex dynamics orders of magnitude faster than conventional solvers, moving digital twins and virtual sensing from offline analysis toward online prediction (Hossain et al., 2025; Kobayashi et al., 2025). Yet their adoption rests on measurement assumptions that operating equipment rarely satisfies.

Existing approaches to learned field prediction fall into two dominant paradigms. The first is autoregressive field rollout, which iteratively predicts the next field frame from the current full-field state, as in graph network simulators, neural weather models (Lam et al., 2023; Bi et al., 2023; Kochkov et al., 2024), message passing PDE solvers (Brandstetter et al., 2022), and latent spectral models (Wu et al., 2023). The second is field-sensor reconstruction, where in-domain sensors measure the same physical quantity as the target field and a learned model recovers the complete field from spatially subsampled measurements, as in sparse-query attention reconstruction (Santos et al., 2023), Voronoi tessellation-assisted field recovery (Fukami et al., 2021), and proxy-based reconstruction

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.

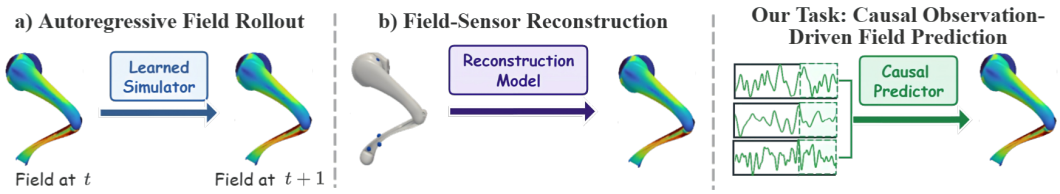

Figure 1: Prediction paradigms. Panel a shows autoregressive rollout advancing the previous full field; panel b shows field-sensor reconstruction inverting in-situ measurements of the same physical quantity. Our task predicts the next-step field solely from state sequences sharing neither units nor locations with the field, with field observations prohibited.

for engineering digital twins (Kobayashi et al., 2025). Both paradigms share one idealized assumption: continuous access to complete full-field states or to in-domain measurements matched to the target field in physical units and spatial support.

In real industrial deployment, this assumption is generally untenable. Dense full-domain instrumentation on complex three-dimensional structures incurs excessive cost, structural intrusion, and maintenance burdens (Sun et al., 2024). Practical systems instead provide only a few system-level state observations, such as actuator cylinder pressures, end-effector loads, and boundary driving velocities (Ruderman et al., 2019; Kružić et al., 2021). These low-dimensional heterogeneous time series differ from the target fields in physical units and spatial distribution, and correspond to no mesh nodes. Autoregressive rollout thus loses its required full-field inputs, and the spatial interpolation bias of in-situ reconstruction no longer applies. We therefore study an underexplored yet practically urgent task: predicting the next-step full physical field on unstructured meshes purely from short windows of heterogeneous state sequences, without any historical or instantaneous field observations. Two open questions arise: to what extent can low-dimensional heterogeneous observations constrain high-dimensional field evolution, and what structural form should an optimal predictor take to approach the inherent observability limit.

We answer these questions theoretically and propose the Sensor-Conditioned Neural Operator with Adaptive Spectral Rotation, termed SONAR. Building on Duhamel-Volterra integral theory and system observability analysis, we prove that the predictable field components are spanned by causal convolutions of observation sequences, that unobservable field energy induces an error floor independent of model capacity, and that the optimal field predictor admits a conditional spectral decomposition. Guided by this theory, SONAR factorizes field prediction into observation-adaptive spectral basis generation and sensor-driven field synthesis. Deep message passing on unstructured meshes produces hierarchical spectral basis fields, which are rotated according to the encoded observation history and orthogonalized through embedded Gram whitening. A gated observation encoder produces basis coefficients for global field synthesis, while a curvature-guided active-set residual branch recovers localized high-frequency details missed by low-rank global reconstruction. Unlike fixed-basis neural operators, SONAR introduces state-conditioned spectral adaptivity, approaching the observability ceiling imposed by sparse heterogeneous sensing.

Our contributions are threefold:

- We formulate an observation-driven causal next-step full-field prediction task on unstructured meshes and construct multi-physics benchmarks covering elastic-beam, heat-plate, and mechanical-leg scenarios.
- We establish an observability and information-budget theory for heterogeneous observation-driven field prediction, characterizing the observation-spanned predictable subspace, the error floor induced by unobservable field energy, and the conditional spectral decomposition of optimal predictors, all verified by ablation, uncertainty, and robustness experiments.
- We propose SONAR, a conditional spectral neural operator that couples observation-rotated Gram-whitened spectral bases with sensor-adaptive coefficient synthesis. Under identical information budgets, SONAR consistently outperforms mainstream baselines on all benchmarks.

2 Related Work

Autoregressive field prediction on meshes. Graph network simulators introduced encode-process-decode message passing for particle-based and mesh-based dynamics (Sanchez-Gonzalez et al., 2020; Pfaff et al., 2021), and attention-based predictors have since scaled the same rollout recipe (Cao et al., 2023; Zhdanov et al., 2025; Hao et al., 2023; Li et al., 2023; Wu et al., 2024; Alkin et al., 2024; Wen et al., 2025), up to learned weather forecasting (Lam et al., 2023; Kochkov et al., 2024) and message passing PDE solvers (Brandstetter et al., 2022). All share one input contract: prediction starts from the previous full field and proceeds autoregressively, so without field observations they cannot even be initialized—precisely the situation our task imposes.

Field reconstruction from sensors. A complementary line recovers full fields from sparse sensors: recurrent decoders of lagged histories (Williams et al., 2024), query-based attention (Santos et al., 2023), Voronoi tessellation (Fukami et al., 2021), wall-measurement inference of near-wall turbulence (Guastoni et al., 2021; Güemes et al., 2021), and proxies-to-fields reconstruction for engineering digital twins (Kobayashi et al., 2025; Hossain et al., 2025). All assume sensors measure the same physical quantity as the target field at in-domain locations—a spatial interpolation or inversion problem—whereas our setting provides only heterogeneous system-level observed state signals sharing neither units nor locations with the field.

Fixed-basis reduced-order modeling. A third family compresses the field onto an offline spatial basis and learns only its coefficients: proper orthogonal decomposition (Sirovich, 1987; Berkooz et al., 1993) underpins model reduction and modal analysis (Benner et al., 2015; Taira et al., 2017), with coefficients regressed by Gaussian processes (Guo & Hesthaven, 2018) or recurrent networks (Williams et al., 2024). Neural operators adopt the same bilinear structure: DeepONet (Lu et al., 2021) pairs a branch encoding the input function with a fixed-basis trunk, and its sequential extensions encode time-dependent loading histories (He et al., 2024c;b;a). A frozen basis, however, represents only a distribution-averaged subspace and cannot adapt modal directions to an individual instance—the structural bottleneck that state-conditioned bases remove.

3 Method

3.1 Problem Formulation

Let $\mathcal{G} = (\mathbb{X}, \mathcal{E})$ be a mesh graph with node coordinates $\mathbb{X} = \{\boldsymbol{x}_i\}_{i=1}^{N}$, $\boldsymbol{x}_i \in \mathbb{R}^{d_x}$, and edge set $\mathcal{E}$, discretizing a physical domain into N nodes. The system is monitored through D_s system-level sensor channels that record observed state signals, of which a causal window $\boldsymbol{S} = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_T\} \in \mathbb{R}^{T \times D_s}$ is available, with $\boldsymbol{s}_t \in \mathbb{R}^{D_s}$ the observed state vector at time t. The observed state channels are heterogeneous in physical units and carry no spatial correspondence with the nodes of $\mathcal{G}$. The target is the full physical field at the next time step, $\boldsymbol{Y} \in \mathbb{R}^{N \times C}$, where C is the number of physical channels per node. The learning problem is

$$\min_{f \in \mathbb{F}} \mathbb{E} \left\| f(\boldsymbol{S}, \mathcal{G}) - \boldsymbol{Y} \right\|_F^2, \tag{1}$$

over a hypothesis class $\mathbb{F}$ of measurable predictors, under the contract that f never accesses the target field or any of its history. By Proposition 1 and Remark 1, this contract removes no information the sensor channel could deliver: the predictable component of $\boldsymbol{Y}$ is exhausted by causal convolutions of $\boldsymbol{S}$, and by Proposition 2 every predictor in $\mathbb{F}$ faces the same error floor—a property of the observation channel, not the predictor—set by the energy of the field outside the predictable subspace.

3.2 Overview

SONAR factorizes the prediction into a conditional spectral synthesis and a localized residual correction,

$$\hat{\boldsymbol{Y}} = \underbrace{\boldsymbol{\Phi}_\theta(\mathcal{G}; \boldsymbol{z}_s)\, \boldsymbol{a}_\psi(\boldsymbol{z}_s)}_{\text{conditional spectral synthesis}} + \underbrace{R_\omega\big(\Omega_a;\, \boldsymbol{z}_s, \mathbb{X}, \hat{\boldsymbol{Y}}_{\text{glob}}\big)}_{\text{active-set residual}}, \tag{2}$$

where $\boldsymbol{z}_s \in \mathbb{R}^{d_h}$ is an embedding of the observation window, $\boldsymbol{\Phi}_\theta(\mathcal{G}; \boldsymbol{z}_s) \in \mathbb{R}^{N \times C \times R}$ is a set of R spectral basis fields generated per sample on the mesh, $\boldsymbol{a}_\psi(\boldsymbol{z}_s) \in \mathbb{R}^R$ is the coefficient vector, $\hat{\boldsymbol{Y}}_{\text{glob}}$

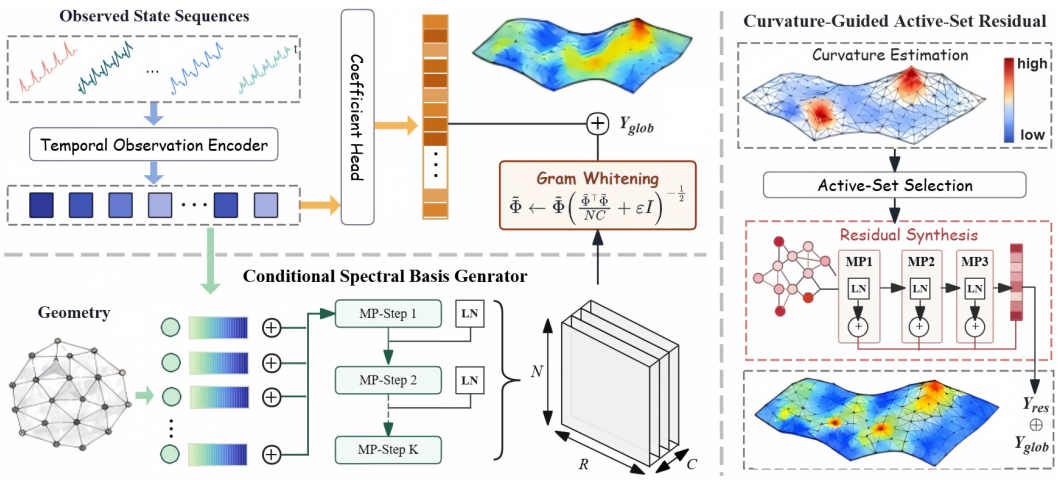

Figure 2: Architecture of SONAR. A temporal encoder compresses observed state sequences into a state embedding. The basis generator maps the mesh and embedding through K message-passing steps into R basis fields of size $N \times C$, orthonormalized by a Gram-whitening layer; a coefficient head produces R coefficients, and their bilinear synthesis forms the global field $\mathbf{Y}_{\text{glob}}$. A curvature-guided active-set branch adds a residual $\mathbf{Y}_{\text{res}}$ on high-curvature nodes.

denotes the bilinear synthesis of the first term, and R_ω is a residual map acting on an active node set $\Omega_a \subseteq \{1, \dots, N\}$. The factorization separates what the system can currently produce, a spatial basis rotated by the current state, from what the input commands, the coordinates of the field within that basis. Proposition 3 shows that a basis frozen across states carries an error term that no amount of additional rank can remove, so the basis must be generated per sample and conditioned on the state. Proposition 7 prescribes where model capacity should live: the geometry-to-basis map is supervised at every node for every basis field and therefore tolerates depth, while the sensor-to-coefficient map outputs only R numbers and should remain shallow. The following subsections instantiate these obligations.

3.3 Temporal Observation Encoder

The observation window is compressed into a state embedding by a gated recurrent unit,

$$\boldsymbol{\gamma}_t = \sigma\big(\boldsymbol{W}_\gamma \boldsymbol{s}_t + \boldsymbol{U}_\gamma \boldsymbol{h}_{t-1} + \boldsymbol{b}_\gamma\big), \quad \boldsymbol{\rho}_t = \sigma\big(\boldsymbol{W}_\rho \boldsymbol{s}_t + \boldsymbol{U}_\rho \boldsymbol{h}_{t-1} + \boldsymbol{b}_\rho\big), \tag{3}$$

$$\tilde{\boldsymbol{h}}_t = \tanh\big(\boldsymbol{W}_h \boldsymbol{s}_t + \boldsymbol{U}_h(\boldsymbol{\rho}_t \odot \boldsymbol{h}_{t-1}) + \boldsymbol{b}_h\big), \qquad \boldsymbol{h}_t = (1 - \boldsymbol{\gamma}_t) \odot \boldsymbol{h}_{t-1} + \boldsymbol{\gamma}_t \odot \tilde{\boldsymbol{h}}_t, \tag{4}$$

for $t = 1, \dots, T$ with $\boldsymbol{h}_0 = \mathbf{0}$, where $\boldsymbol{\gamma}_t, \boldsymbol{\rho}_t \in [0,1]^{d_h}$ are the update and reset gates, σ is the logistic function, $\odot$ denotes elementwise multiplication, and $\boldsymbol{W}_\cdot \in \mathbb{R}^{d_h \times D_s}$, $\boldsymbol{U}_\cdot \in \mathbb{R}^{d_h \times d_h}$, $\boldsymbol{b}_\cdot \in \mathbb{R}^{d_h}$ are learnable parameters. The embedding is the final hidden state, $\boldsymbol{z}_s \triangleq \boldsymbol{h}_T$.

This choice is dictated by the realization theory of Appendix A.7. By Proposition 1, the temporal side of the predictor approximates causal Green kernels of the forced dynamics, and Proposition 8 shows that their canonical realization is a state-space recursion with infinite impulse response, while any short finite impulse response truncation incurs a tail error that diverges as the system poles approach the unit circle. The gated update above is a nonlinear recursion with time-varying effective poles and therefore belongs to the admissible function class, providing exponentially decaying long memory without committing to a fixed kernel length.

3.4 Conditional Spectral Basis Generator

The basis generator maps the mesh and the state embedding to R basis fields and carries the full depth of the model, in accordance with Proposition 7.

Node and edge features. Each node is initialized from its coordinates and the broadcast state embedding, $\boldsymbol{v}_i^0 = [\boldsymbol{x}_i; \boldsymbol{z}_s] \in \mathbb{R}^{d_x + d_h}$, and each edge $(i,j) \in \mathcal{E}$ is described by the unit direction and

length of its segment, $\boldsymbol{e}_{ij} = \big[(\boldsymbol{x}_j - \boldsymbol{x}_i)/\|\boldsymbol{x}_j - \boldsymbol{x}_i\|_2;\ \|\boldsymbol{x}_j - \boldsymbol{x}_i\|_2\big]$. Node features are lifted by an encoder, $\boldsymbol{h}_i^0 = \mathrm{MLP}_{\mathrm{enc}}(\boldsymbol{v}_i^0) \in \mathbb{R}^d$, and edge features are embedded as $\bar{\boldsymbol{e}}_{ij} \in \mathbb{R}^d$.

Deep message passing. For $k = 1, \dots, K$ steps with per-step parameters θ_k that are not shared across steps,

$$\boldsymbol{m}_{ij}^k = \phi_k\big(\boldsymbol{h}_i^{k-1}, \boldsymbol{h}_j^{k-1}, \bar{\boldsymbol{e}}_{ij}\big), \qquad \boldsymbol{h}_i^k = \mathrm{LN}\Big(\boldsymbol{h}_i^{k-1} + \psi_k\big(\boldsymbol{h}_i^{k-1}, \textstyle\sum_{j\in\mathbb{N}(i)} \boldsymbol{m}_{ij}^k\big)\Big), \tag{5}$$

where ϕ_k, ψ_k are two-layer networks, $\mathbb{N}(i)$ is the neighborhood of node i, aggregation is summation, and LN denotes layer normalization around a residual connection. Unshared weights let depth translate directly into the expressivity of the generated basis family, which is the mechanism behind the strict inclusion of Proposition 4, while Proposition 6 guarantees that this family remains dense enough to preserve universal approximation of the underlying operator.

Basis fields. The final representation is decoded per node and per channel into basis fields,

$$\boldsymbol{\Phi}_\theta = \mathrm{MLP}_{\mathrm{basis}}\big(\boldsymbol{H}^K\big) \in \mathbb{R}^{N\times C\times R}, \qquad \Phi_{i,c,r} = [\boldsymbol{\phi}_r]_{i,c}, \tag{6}$$

so that $\boldsymbol{\phi}_r \in \mathbb{R}^{N\times C}$ is the r-th basis field. Per-channel bases allow different physical components to own distinct spatial modes.

Gram whitening as a structural component. Flattening $\boldsymbol{\Phi}$ to $\bar{\boldsymbol{\Phi}} \in \mathbb{R}^{NC\times R}$, we apply one differentiable whitening step per sample,

$$\bar{\boldsymbol{\Phi}} \ \leftarrow\ \bar{\boldsymbol{\Phi}}\big(\bar{\boldsymbol{\Phi}}^\top\bar{\boldsymbol{\Phi}}/(NC) + \epsilon\boldsymbol{I}\big)^{-1/2}, \tag{7}$$

implemented through a Cholesky factorization with a small jitter ϵ. Proposition 5 gives the reason this layer exists: the bilinear model is invariant under any invertible reparameterization of the basis, so coefficients are unidentifiable and unevenly scaled unless the continuous symmetry is quotiented out; whitening restricts the residual gauge freedom to signed permutations and makes the coefficient-side curvature nearly isotropic. Crucially, orthogonality is enforced architecturally by Eq. equation 7 and never appears as a loss term, which keeps the training objective free of competing penalties.

3.5 Coefficient Head and Global Synthesis

The coefficients are produced by a shallow network from the state embedding and a pooled geometry embedding,

$$\boldsymbol{a}_\psi = \mathrm{MLP}_{\mathrm{coef}}\big([\boldsymbol{z}_s; \boldsymbol{z}_G]\big) \in \mathbb{R}^R, \qquad \boldsymbol{z}_G = \mathrm{Pool}\big(\{\boldsymbol{h}_i^K\}_{i=1}^N\big), \tag{8}$$

where $\boldsymbol{z}_G$ is a permutation-invariant global pooling of the mesh representation, retained so that the same interface supports meshes that vary across samples. The head is deliberately shallow because its output is only R-dimensional, placing it on the supervision-sparse side identified by Proposition 7. The global field is then synthesized channelwise,

$$[\hat{\boldsymbol{Y}}_{\mathrm{glob}}]_{i,c} = \sum_{r=1}^{R} a_r\, \Phi_{i,c,r}. \tag{9}$$

Equation equation 9 is the branch–trunk form of Lu et al. (2021) lifted to meshes, with two essential differences: the trunk is generated per sample by Eq. equation 5 rather than fixed after training, and it is orthonormalized by Eq. equation 7. Proposition 4 shows that this strictly enlarges the fixed-basis hypothesis class and that warm-starting the generator at a frozen basis cannot increase the training objective under full-batch descent, with an expected gain equal to the across-state variation removed by conditioning.

3.6 Curvature-Guided Active-Set Residual

The synthesis in Eq. equation 9 has rank at most R and therefore cannot express localized high-frequency structure such as stress concentrations or shear layers. The residual branch concentrates a small correction capacity exactly where such structure appears.

Curvature estimate. For the detached global field, a discrete Laplacian energy serves as a curvature proxy per node,

$$\hat{\kappa}_i = \sum_{c=1}^{C} \Big(\sum_{j \in \mathbb{N}(i)} w_{ij} \big([\hat{\boldsymbol{Y}}_{\text{glob}}]_{j,c} - [\hat{\boldsymbol{Y}}_{\text{glob}}]_{i,c} \big) \Big)^2, \qquad w_{ij} = \frac{1}{|\mathbb{N}(i)|}, \tag{10}$$

which captures local second-order variation without additional supervision.

Active set. The active set is seeded by the nodes with the highest curvature, at a fixed quantile q of $\{\hat{\kappa}_i\}$, and expanded by h hops along the graph topology,

$$\Omega_a = \text{Hop}_h\big(\{ i : \hat{\kappa}_i \geq Q_q(\hat{\kappa}) \} \big), \tag{11}$$

so that the receptive field of the residual covers the full support of each curvature peak.

Residual synthesis. On the subgraph induced by Ω_a, a shallow message-passing stack with node features $[\boldsymbol{x}_i; \boldsymbol{z}_s; [\hat{\boldsymbol{Y}}_{\text{glob}}]_{i,:}]$ produces a correction that is zero-padded back to the full mesh,

$$\hat{\boldsymbol{Y}} = \hat{\boldsymbol{Y}}_{\text{glob}} + \text{Pad}\Big(\text{MLP}_{\text{dec}}\big(\text{MP}_{\Omega_a}([\mathbb{X}, \boldsymbol{z}_s, \hat{\boldsymbol{Y}}_{\text{glob}}]) \big) \Big). \tag{12}$$

Computing Eq. equation 12 only on Ω_a confines the cost of the correction to the small fraction of nodes that need it, while the global structure remains entirely in charge of the basis synthesis.

3.7 Training Objective

The model is trained end to end with a single objective in the normalized field space,

$$\mathbb{L} = \underbrace{\frac{\|\hat{\boldsymbol{Y}} - \boldsymbol{Y}\|_F^2}{\|\boldsymbol{Y}\|_F^2}}_{\text{normalized squared error}} + \lambda \underbrace{\frac{\|\hat{\boldsymbol{Y}} - \boldsymbol{Y}\|_1}{\|\boldsymbol{Y}\|_1}}_{\text{normalized absolute error}}, \qquad \lambda < 1, \tag{13}$$

where $\boldsymbol{Y}$ is standardized with training-set statistics. The squared term dominates the fit of the global field energy, while the absolute term at reduced weight suppresses the heavy tail of large elementwise errors. No other penalty is added: all structural constraints, in particular the orthonormality of the basis, are realized architecturally rather than as terms in Eq. equation 13, so the optimization landscape is shaped by construction instead of by competing losses. experiments

4 Experiments

4.1 Experimental Setup

Datasets. We evaluate on three synthetic benchmarks. ElasticBeam-Syn is a linear-elastic cantilever beam with a placed hole under randomized loads, targeting the six-component stress field. HeatPlate-Syn is a transient heat-conduction plate with randomized sources and boundary conditions, targeting the three-component heat-flux field. Go1Leg-Dyn is a linear-elastodynamic stress field on the front leg of a Unitree Go1 robot driven by randomized joint torques propagated through MuJoCo (Todorov et al., 2012), with a seventeen-dimensional system-level sensor input. Within each dataset, geometry and materials are fixed and the observed state sequence is the only per-sample variation. Full details, including sample counts and splits, are given in Appendix B.

Baselines. We compare against nine representative methods in three families. The first family is sequence-to-field operator models that ignore mesh topology: DeepONet (Lu et al., 2021), its sequential extension S-DeepONet (He et al., 2024c), and the shallow recurrent decoder SHRED (Williams et al., 2024). The second family is mesh graph neural simulators: MeshGraphNets (Pfaff et al., 2021), its multi-scale variant BSMS-GNN (Cao et al., 2023), the hierarchical transformer Erwin (Zhdanov et al., 2025), and the recurrent graph model LSTM-GNN (Guevara Garban et al., 2026). The third family is attention-based field predictors: GNOT (Hao et al., 2023) and Transolver (Wu et al., 2024).

Table 1: Main comparison on the three benchmarks. Entries are means over five training seeds; **bold** and underline mark the best and second best per column. Improvement is SONAR's relative reduction against the strongest baseline, and p-values report Welch's t-tests across seeds. Metric definitions are in Appendix C; the full table with standard deviations is in Appendix E.

	ElasticBeam-Syn			HeatPlate-Syn			Go1Leg-Dyn		
Method	NMSE↓	RelL2↓	VM-MaxRE↓	NMSE↓	RelL2↓	VM-MaxRE↓	NMSE↓	RelL2↓	VM-MaxRE↓
DeepONet	0.3889	0.3048	4.549	0.1373	0.2989	13.464	0.8899	0.7963	346
MeshGraphNets	0.0263	0.1606	2.851	0.0007	0.0318	0.784	0.2390	0.4017	26.4
BSMS-GNN	0.0469	0.1853	3.745	0.0046	0.0562	1.577	0.3094	0.4864	32.4
GNOT	0.1113	0.2524	4.166	0.0105	0.0762	3.001	0.5823	0.5444	102
SHRED	0.0397	0.2187	3.043	0.0023	0.0354	0.763	0.2522	0.4739	13.3
S-DeepONet	0.3879	0.3010	4.370	0.1275	0.2774	12.965	0.8393	0.7094	372
Transolver	0.3625	0.2816	5.545	0.0252	0.1273	7.303	0.4438	0.4876	87.5
Erwin	0.0506	0.2036	3.863	0.0056	0.0570	1.792	0.3398	0.4537	169
LSTM-GNN	0.0604	0.2514	3.939	0.0007	0.0333	0.742	0.2122	0.3930	43.9
SONAR (Ours)	**0.0249**	**0.1600**	**2.730**	**0.0005**	**0.0281**	**0.734**	**0.1897**	**0.3580**	**11.7**
Improvement	5.3%	0.4%	4.2%	28.6%	11.6%	1.1%	10.6%	8.9%	12.0%
p-value	0.043	0.872	0.037	0.004	<0.001	0.054	<0.001	0.004	0.003

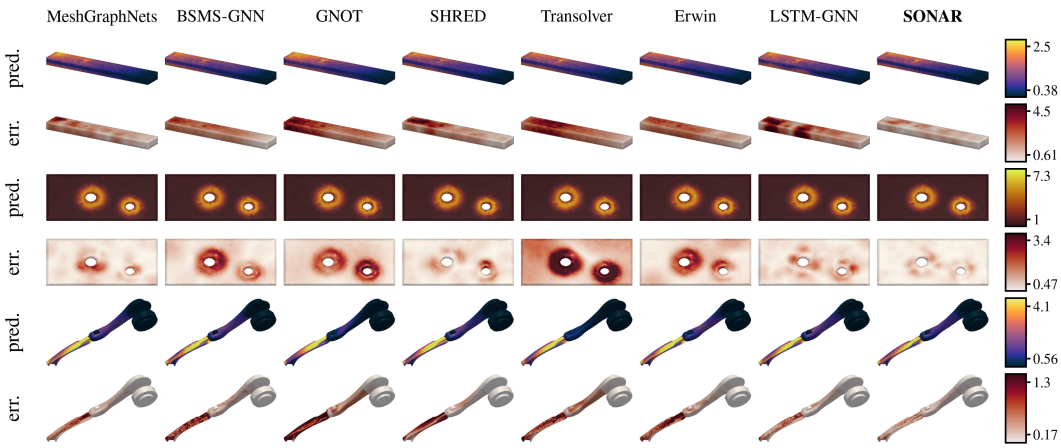

Figure 3: Predicted fields and error maps of representative methods on the three benchmarks.

Implementation details. All methods share the same data split, normalization, and training objective of Eq. equation 13, and all results are averaged over five training seeds with the split fixed throughout. Each baseline is tuned by grid search on the validation split and reported at its best validated configuration; the unified protocol and search ranges are given in Appendix D. Methods that natively require field history have their field input channels removed and replaced by the broadcast sensor embedding, so that all methods operate under the identical information budget; sensor-conditioned baselines share the gated recurrent encoder of SONAR, except SHRED and LSTM-GNN, which retain their native recurrent encoders at matched hidden width. SONAR uses $K = 15$ message-passing steps and $R = 64$ basis fields.

4.2 Comparison with Baselines

Table 1 and Figure 3 report the main comparison across the three datasets. SONAR achieves the best mean on all nine dataset-metric combinations, significantly so in seven of nine ($p < 0.05$). The global-metric margin is largest on the spectrally concentrated HeatPlate-Syn, while the tail-error gain is largest on Go1Leg-Dyn, the benchmark with the most complex geometry and the slowest spectral decay; on ElasticBeam-Syn the gains on NMSE and VM-MaxRE are significant, with RelL2 statistically tied.

The per-family comparison clarifies the source of these gains. Sequence-to-field operator models with frozen trunk bases (DeepONet, S-DeepONet) fail by large margins on all three benchmarks:

Table 2: Component ablation on ElasticBeam-Syn and Go1Leg-Dyn. Each row removes one component of SONAR. The full model is shaded and the worst entry per column is marked with †.

	ElasticBeam-Syn			Go1Leg-Dyn		
Variant	NMSE↓	RelL2↓	VM-MaxRE↓	NMSE↓	RelL2↓	VM-MaxRE↓
SONAR full	**0.0249** ±0.0009	**0.1600** ±0.0052	**2.730** ±0.07	**0.1897** ±0.0040	**0.3580** ±0.0126	**11.7** ±0.6
w/o state conditioning	0.0612 ±0.0034†	0.2481 ±0.0058†	3.98 ±0.21†	0.2984 ±0.0101†	0.5123 ±0.0097†	28.4 ±1.8†
w/o Gram whitening	0.0315 ±0.0042	0.1783 ±0.0061	3.12 ±0.28	0.2193 ±0.0148	0.3882 ±0.0113	14.9 ±1.5
w/o residual branch	0.0278 ±0.0015	0.1687 ±0.0028	3.85 ±0.17	0.2012 ±0.0070	0.3684 ±0.0061	19.6 ±1.1
w/o curvature guidance	0.0263 ±0.0013	0.1642 ±0.0024	3.24 ±0.14	0.1951 ±0.0066	0.3627 ±0.0058	15.8 ±0.9
shared MP weights	0.0289 ±0.0018	0.1739 ±0.0033	3.31 ±0.16	0.2261 ±0.0089	0.3914 ±0.0072	15.2 ±1.0
w/o geometry pooling	0.0251 ±0.0012	0.1608 ±0.0022	2.77 ±0.10	0.1904 ±0.0064	0.3589 ±0.0056	11.9 ±0.6

they cannot follow the per-sample rotation of modal directions, the directional error term of Proposition 3. SHRED isolates this term: its frozen POD basis leaves only a small marginal projection deficit (Remark 3), so its residual gap to SONAR measures the across-state directional variation that a per-instance basis recovers. Attention-based predictors improve only modestly, and mesh graph simulators form the strongest family, consistent with deep message passing implicitly constructing a per-sample spatial representation; SONAR makes this representation explicit, orthonormalized, and factorized into a state-rotated basis and sensor-driven coefficients, strictly enlarging the fixed-basis hypothesis class (Proposition 4). Finally, the best VM-MaxRE on every dataset shows that the gains extend to the tail errors that dominate structural safety, and Figure 3 confirms this visually with the faintest and most localized error maps.

4.3 Ablation Studies

We ablate SONAR on ElasticBeam-Syn and Go1Leg-Dyn under the split and protocol of Table 1. Table 2 removes one component at a time, Figure 4 sweeps the structural hyperparameters, and Table 3 replaces the temporal encoder. All entries report means over five training seeds, with compact standard deviations shown next to the corresponding means.

Components. State conditioning is the largest contributor: freezing the basis at an equal-rank training-set POD gives the worst entry in every column of Table 2, as predicted by Propositions 3 and 4 and decomposed in Remark 3. Gram whitening contributes differently: removing it raises NMSE only moderately but inflates the seed standard deviation severalfold—exactly the signature predicted by Proposition 5, since unidentifiable coefficients make training sensitive to initialization. The residual branch shows the complementary pattern: its removal costs little NMSE yet raises VM-MaxRE sharply, confirming that the low-rank synthesis captures the global field energy while the active-set residual repairs the localized tail that governs safety margins; a random active set recovers only about half of this gain, so curvature guidance rather than added capacity is responsible.

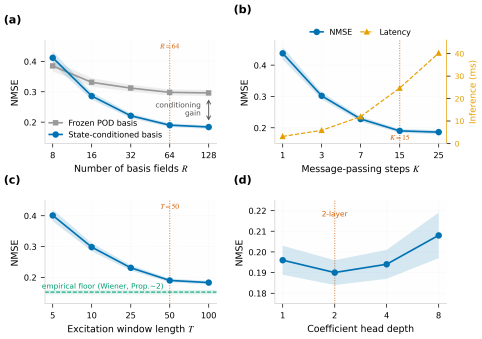

Figure 4: Structural sweeps of R, K, T, and coefficient-head depth on Go1Leg-Dyn.

Structure and hyperparameters. Figure 4 reports the structural sweeps used to select the default configuration. The number of basis fields R improves the conditioned basis monotonically while the frozen basis saturates, so additional rank cannot substitute for directional adaptivity. Accuracy saturates at $K{=}15$ message-passing steps while latency grows linearly, fixing the default at the knee of the trade-off. Lengthening the observation window T lowers the error toward the empirical observability floor estimated by the Wiener solution without crossing it, consistent with Proposition 2 and Remark 2. Finally, a two-layer coefficient head is optimal and deeper heads degrade accuracy, matching the capacity allocation predicted by Proposition 7.

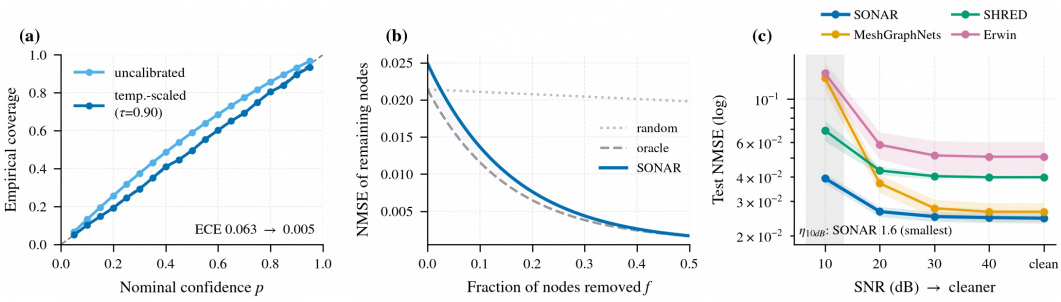

Figure 5: Uncertainty and robustness of SONAR on ElasticBeam-Syn. Left: reliability diagram of the seed ensemble before and after temperature scaling. Middle: sparsification curves with oracle and random references. Right: test NMSE under Gaussian sensor noise; the shaded band marks the severe noise regime.

Temporal encoder. Table 3 replaces only the observation encoder. GRU and LSTM differ by 2% NMSE, so the choice among gated recurrent cells is a second-order effect. Finite impulse response encoders degrade sharply and the flattened window MLP reaches 79% higher NMSE than the GRU. Truncated kernels cannot represent the slowly decaying memory of elastic systems, as Proposition 8 and Remark 5 predict, and the gap widens as the kernel shortens.

Table 3: Temporal encoder ablation on Go1Leg-Dyn. Only the observation encoder is replaced.

Family	Encoder	NMSE↓	RelL2↓	VM-MaxRE↓
IIR	GRU, ours	**0.1897** ±0.0040	**0.3580** ±0.0126	**11.7** ±0.6
IIR	LSTM	0.1934 ±0.0071	0.3627 ±0.0062	12.1 ±0.7
IIR	Transformer	0.2015 ±0.0088	0.3712 ±0.0074	12.8 ±0.8
FIR	Conv. kernel 50	0.2618 ±0.0104	0.4389 ±0.0089	17.3 ±1.2
FIR	Conv. kernel 20	0.3142 ±0.0136	0.4801 ±0.0112	21.5 ±1.6
FIR	MLP flat window	0.3397 ±0.0151	0.5023 ±0.0128	24.8 ±1.9

4.4 Uncertainty Analysis

We equip SONAR with predictive uncertainty via a deep ensemble of the five training seeds of Section 4.1, whose per-node prediction standard deviation σ serves as the uncertainty field at no additional training cost; Figure 5 summarizes three tests on ElasticBeam-Syn. First, σ is quantitatively trustworthy: a single temperature parameter fitted on the validation set reduces the expected calibration error from 0.063 to 0.005, after which the coverage curve nearly coincides with the diagonal. Second, σ localizes failure: removing nodes in decreasing order of σ tracks the oracle ordering by true error and lies far below random removal, and the nodes removed first concentrate at the hole-edge stress concentration—the spatial signature of the unobservable field energy predicted by Proposition 2. Third, the model is robust to corrupted observations: under Gaussian sensor noise down to 10 dB SNR, SONAR attains the lowest test NMSE at every noise level with only $1.6\times$ degradation relative to the clean error, indicating that the whitened basis and the shallow coefficient head do not amplify input noise—essential for deployment on real sensor channels.

5 Limitations

All benchmarks are synthetically generated, and each dataset uses a single fixed geometry; although the pooled geometry embedding allows the interface to support varying meshes, generalization across geometries, and transfer to real deployments with unmodeled physics and imperfect sensors, remains to be validated.

6 Conclusion

We studied causal next-step full-field prediction from short observed state sequences with field observations excluded. Theory shows the predictable field is exhausted by causal convolutions of the observed states and admits a conditional spectral decomposition; SONAR instantiates it with state-rotated, whitened spectral bases, attaining the best accuracy on three benchmarks, and pinpoints where prediction is fundamentally impossible.

REFERENCES

Benedikt Alkin, Andreas Fürst, Simon Steffen Schmid, Lukas Gruber, Markus Holzleitner, and Johannes Brandstetter. Universal physics transformers: A framework for efficiently scaling neural operators. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

Peter Benner, Serkan Gugercin, and Karen Willcox. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM Review*, 57(4):483–531, 2015.

Gal Berkooz, Philip Holmes, and John L. Lumley. The proper orthogonal decomposition in the analysis of turbulent flows. *Annual Review of Fluid Mechanics*, 25:539–575, 1993.

Kaifeng Bi, Lingxi Xie, Hengheng Zhang, Xin Chen, Xiaotao Gu, and Qi Tian. Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, 619(7970):533–538, 2023.

Johannes Brandstetter, Daniel Worrall, and Max Welling. Message passing neural PDE solvers. In *International Conference on Learning Representations (ICLR)*, 2022.

Yadi Cao, Menglei Chai, Minchen Li, and Chenfanfu Jiang. Efficient learning of mesh-based physical simulation with bi-stride multi-scale graph neural network. In *International Conference on Machine Learning (ICML)*, 2023.

Kai Fukami, Romit Maulik, Nesar Ramachandra, Koji Fukagata, and Kunihiko Taira. Global field reconstruction from sparse sensors with Voronoi tessellation-assisted deep learning. *Nature Machine Intelligence*, 3(11):945–951, 2021.

Luca Guastoni, Alejandro Güemes, Andrea Ianiro, Stefano Discetti, Philipp Schlatter, Hossein Azizpour, and Ricardo Vinuesa. Convolutional-network models to predict wall-bounded turbulence from wall quantities. *Journal of Fluid Mechanics*, 928:A27, 2021.

Alejandro Güemes, Stefano Discetti, Andrea Ianiro, Beril Sirmacek, Hossein Azizpour, and Ricardo Vinuesa. From coarse wall measurements to turbulent velocity fields through deep learning. *Physics of Fluids*, 33(7):075121, 2021.

Manuel Ricardo Guevara Garban, Yves Chemisky, Étienne Prulière, Michaël Clément, Martin Abendroth, and Björn Kiefer. Non-linear mechanical field reconstruction coupling recurrent neural networks with physics-informed graph neural networks. *arXiv preprint arXiv:2606.10909*, 2026.

Mengwu Guo and Jan S. Hesthaven. Reduced order modeling for nonlinear structural analysis using Gaussian process regression. *Computer Methods in Applied Mechanics and Engineering*, 341: 807–826, 2018.

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 *International Conference on Machine Learning (ICML)*, pp. 12556–12569, 2023.

Junyan He, Seid Koric, Diab Abueidda, Ali Najafi, and Iwona Jasiuk. Geom-DeepONet: A point-cloud-based deep operator network for field predictions on 3D parameterized geometries. *Computer Methods in Applied Mechanics and Engineering*, 429:117130, 2024a.

Junyan He, Shashank Kushwaha, Jaewan Park, Seid Koric, Diab Abueidda, and Iwona Jasiuk. Predictions of transient vector solution fields with sequential deep operator network. *Acta Mechanica*, 235(8):5257–5272, 2024b.

Junyan He, Shashank Kushwaha, Jaewan Park, Seid Koric, Diab Abueidda, and Iwona Jasiuk. Sequential deep operator networks (S-DeepONet) for predicting full-field solutions under time-dependent loads. *Engineering Applications of Artificial Intelligence*, 127:107258, 2024c.

B. L. Ho and Rudolf E. Kálmán. Effective construction of linear state-variable models from input/output functions. *Regelungstechnik*, 14(12):545–548, 1966.

Raisa Hossain, Farid Ahmed, Kazuma Kobayashi, Seid Koric, Diab Abueidda, and Syed Bahauddin Alam. Virtual sensing-enabled digital twin framework for real-time monitoring of nuclear systems leveraging deep neural operators. *npj Materials Degradation*, 9(1):21, 2025.

Rudolf E. Kalman. Mathematical description of linear dynamical systems. *SIAM Journal on Control*, 1(2):152–192, 1963.

Kazuma Kobayashi, Samrendra Roy, Seid Koric, Diab Abueidda, and Syed Bahauddin Alam. From proxies to fields: Spatiotemporal reconstruction of global radiation from sparse sensor sequences. *arXiv preprint arXiv:2506.12045*, 2025.

Dmitrii Kochkov, Janni Yuval, Ian Langmore, Peter Norgaard, Jamie Smith, Griffin Mooers, Milan Klöwer, James Lottes, Stephan Rasp, Peter Düben, Sam Hatfield, Peter Battaglia, Alvaro Sanchez-Gonzalez, Matthew Willson, Michael P. Brenner, and Stephan Hoyer. Neural general circulation models for weather and climate. *Nature*, 632(8024):1060–1066, 2024.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to PDEs. *Journal of Machine Learning Research*, 24(89):1–97, 2023.

Stanko Kružić, Josip Musić, Roman Kamnik, and Vladan Papić. End-effector force and joint torque estimation of a 7-dof robotic manipulator using deep learning. *Electronics*, 10(23):2963, 2021.

Remi Lam, Alvaro Sanchez-Gonzalez, Matthew Willson, Peter Wirnsberger, Meire Fortunato, Ferran Alet, Suman Ravuri, Timo Ewalds, Zach Eaton-Rosen, Weihua Hu, Alexander Merose, Stephan Hoyer, Dominic King, Oriol Vinyals, Tim Stott, Alexander Pritzel, Shakir Mohamed, and Peter Battaglia. Learning skillful medium-range global weather forecasting. *Science*, 382 (6677):1416–1421, 2023.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations (ICLR)*, 2021.

Zongyi Li, Nikola Kovachki, Chris Choy, Boyi Li, Jean Kossaifi, Shourya Otta, Mohammad Amin Nabian, Maximilian Stadler, Christian Hundt, Kamyar Azizzadenesheli, and Anima Anandkumar. Geometry-informed neural operator for large-scale 3D PDEs. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, 2021.

Tobias Pfaff, Meire Fortunato, Alvaro Sanchez-Gonzalez, and Peter Battaglia. Learning mesh-based simulation with graph networks. In *International Conference on Learning Representations (ICLR)*, 2021.

Michael Ruderman, Leonid Fridman, and Philipp Pasolli. Virtual sensing of load forces in hydraulic actuators using second- and higher-order sliding modes. *Control Engineering Practice*, 92:104151, 2019.

Alvaro Sanchez-Gonzalez, Jonathan Godwin, Tobias Pfaff, Rex Ying, Jure Leskovec, and Peter Battaglia. Learning to simulate complex physics with graph networks. In *International Conference on Machine Learning (ICML)*, 2020.

Jesus E. Santos, Zachary R. Fox, Arvind Mohan, Daniel O'Malley, Hari Viswanathan, and Nicholas Lubbers. Development of the Senseiver for efficient field reconstruction from sparse observations. *Nature Machine Intelligence*, 5(11):1317–1325, 2023.

Lawrence Sirovich. Turbulence and the dynamics of coherent structures. Part I: Coherent structures. *Quarterly of Applied Mathematics*, 45(3):561–571, 1987.

L. Sun, H. Sun, W. Zhang, and Y. Li. Hybrid monitoring methodology: A model-data integrated digital twin framework for structural health monitoring and full-field virtual sensing. *Advanced Engineering Informatics*, 60:102386, 2024.

Kunihiko Taira, Steven L. Brunton, Scott T. M. Dawson, Clarence W. Rowley, Tim Colonius, Beverley J. McKeon, Oliver T. Schmidt, Stanislav Gordeyev, Vassilios Theofilis, and Lawrence S. Ukeiley. Modal analysis of fluid flows: An overview. *AIAA Journal*, 55(12):4013–4041, 2017.

Emanuel Todorov, Tom Erez, and Yuval Tassa. MuJoCo: A physics engine for model-based control. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5026–5033, 2012.

Shizheng Wen, Arsh Kumbhat, Levi 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. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

Jan P. Williams, Olivia Zahn, and J. Nathan Kutz. Sensing with shallow recurrent decoder networks. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 480(2298):20240054, 2024.

Haixu Wu, Tengge Hu, Huakun Luo, Jianmin Wang, and Mingsheng Long. Solving high-dimensional PDEs with latent spectral models. In *International Conference on Machine Learning (ICML)*, 2023.

Haixu Wu, Huakun Luo, Haowen Wang, Jianmin Wang, and Mingsheng Long. Transolver: A fast transformer solver for PDEs on general geometries. In *International Conference on Machine Learning (ICML)*, 2024.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations (ICLR)*, 2019.

Maksim Zhdanov, Max Welling, and Jan-Willem van de Meent. Erwin: A tree-based hierarchical transformer for large-scale physical systems. In *International Conference on Machine Learning (ICML)*, 2025.

A Theoretical Framework

This appendix develops the theoretical backbone of the paper. Throughout we follow the notation convention: italic letters for scalars, bold lowercase letters for vectors, bold uppercase letters for matrices, bold sans-serif uppercase letters for tensors, and script letters for sets and graphs; Table 4 summarizes every symbol used in the proofs.

A.1 Problem Setup, Assumptions, and Notation

Task. A physical field defined on a mesh graph $\mathcal{G} = (\mathbb{X}, \mathcal{E})$ with N nodes and C channels per node is sampled as a state vector $\boldsymbol{u}(t) \in \mathbb{R}^{NC}$. The predictor receives a window of observed state sequences $\boldsymbol{S} = \{\boldsymbol{s}_1, \dots, \boldsymbol{s}_T\} \in \mathbb{R}^{T \times D_s}$, $\boldsymbol{s}_t \in \mathbb{R}^{D_s}$ (partial observations of related physical quantities of the same system), together with $\mathcal{G}$, and outputs a prediction $\hat{\boldsymbol{u}}$ of the next-step field $\boldsymbol{u}(T{+}1)$. The information budget of the input is $M \triangleq T \cdot D_s$ scalars. The predictor is any measurable map $f : (\boldsymbol{S}, \mathcal{G}) \mapsto \hat{\boldsymbol{u}}$; it may not access $\boldsymbol{u}(\tau)$ or any function thereof for any $\tau \leq T$.

Assumption 1 (Forced dynamics)**.** *The field evolves according to $\dot{\boldsymbol{u}}(t) = \boldsymbol{f}(\boldsymbol{u}(t), \boldsymbol{s}(t))$ with $\boldsymbol{f}$ smooth, and a unique solution exists on the window. Linearized about a reference state,*

$$\dot{\boldsymbol{u}}(t) = \boldsymbol{A}\boldsymbol{u}(t) + \boldsymbol{B}\boldsymbol{s}(t) + \boldsymbol{n}(\boldsymbol{u}(t)), \tag{14}$$

where $\boldsymbol{A} \in \mathbb{R}^{NC \times NC}$ is the free-dynamics operator, $\boldsymbol{B} \in \mathbb{R}^{NC \times D_s}$ is the observation-to-field input operator, and $\boldsymbol{n}$ collects analytic higher-order terms with multilinear coefficients $\mathbb{N}_j$, $j \geq 2$.

Assumption 2 (Stability)**.** *$\boldsymbol{A}$ is diagonalizable and Hurwitz; the eigenvalues $\{\lambda_p\}$ of the associated discrete-time propagator satisfy $|\lambda_p| \leq 1 - \varepsilon_0$ for some $\varepsilon_0 > 0$ that may be arbitrarily small.*

Assumption 3 (Orthogonality)**.** *The free-response component driven by the unknown initial state is independent of the observation channel: writing the predictable (forced) subspace as $\mathbb{U}_{\mathrm{obs}}$ (Definition 1) with orthogonal projector $\boldsymbol{P}$, the component $\boldsymbol{P}_\perp \boldsymbol{u} \triangleq (\boldsymbol{I} - \boldsymbol{P})\boldsymbol{u}$ is centered and independent of $(\boldsymbol{S}, \mathcal{G})$, the only information channel available to the predictor under the task contract (see Remark 1 for the scope of this assumption).*

Assumption 4 (Conditional second-order structure)**.** *Conditioned on a state variable g (geometry and operating condition), the centered field has mean $\boldsymbol{\mu}_g$ and covariance $\boldsymbol{C}_g$; the across-condition mean variation $\boldsymbol{M} \triangleq \mathrm{Var}_g[\boldsymbol{\mu}_g]$ has effective rank r_{eff} with a nearly flat spectrum on its support: $\lambda_1(\boldsymbol{M}) \le \big(1/r_{\mathrm{eff}}\big)\,\mathrm{tr}(\boldsymbol{M})\,(1+o(1))$.*

Assumption 5 (Boundedness)**.** *Embeddings and coefficients are bounded: $\|\boldsymbol{z}_s\|_2 \le B_z$, and the coefficient map $\boldsymbol{a} = \boldsymbol{W}\boldsymbol{z}_s$ is constrained by $\|\boldsymbol{W}\|_F \le B_W$; the target satisfies $\|\boldsymbol{a}\|_2 \le B_a$.*

Table 4: Notation used throughout Appendix A.

Symbol	Type	Meaning
N, C, R	scalar	numbers of mesh nodes, field channels, and basis fields
T, D_s, M	scalar	window length, observed-state dimension, budget $M = TD_s$
n	scalar	number of training samples
t, Δt	scalar	time index and step
$\boldsymbol{u}(t)$	vector $\in \mathbb{R}^{NC}$	sampled physical field at time t; $\hat{\boldsymbol{u}}$ its prediction
$\boldsymbol{s}_t$, $\boldsymbol{S}$	vector / matrix	observed state at time t; window $\boldsymbol{S} \in \mathbb{R}^{T\times D_s}$
$\mathcal{G} = (\mathbb{X}, \mathcal{E})$	graph	mesh with node coordinates $\mathbb{X}$ and edges $\mathcal{E}$
$\boldsymbol{A}$, $\boldsymbol{B}$	matrix	free-dynamics and input operators in Eq. equation 14
$\mathbb{N}_j$	j-linear map	order-j nonlinear term of the dynamics
$\mathbf{G}_m$	tensor	order-m Volterra kernel, $\mathbf{G}_1$ the linear Green kernel
$\mathbb{U}_{\mathrm{obs}}$, $\boldsymbol{P}$, $\boldsymbol{P}_\perp$	subspace / matrices	predictable subspace and its projectors
g, $\boldsymbol{\mu}_g$, $\boldsymbol{C}_g$	variable / vector / matrix	condition, conditional mean, conditional covariance
$\boldsymbol{M}$, r_{eff}	matrix / scalar	$\mathrm{Var}_g[\boldsymbol{\mu}_g]$ and its effective rank
$\lambda_i(\cdot)$	function	i-th largest eigenvalue
$\boldsymbol{\Phi}$	matrix $\in \mathbb{R}^{NC\times R}$	basis fields, $\boldsymbol{\Phi}_\theta$ when generated by a network
$\boldsymbol{a}$, $\boldsymbol{z}_s$	vector	basis coefficients; sensor embedding $\boldsymbol{z}_s \in \mathbb{R}^{d_h}$
$\boldsymbol{W}$	matrix	invertible gauge map, or coefficient-head weight
λ_p, $\boldsymbol{c}_p$	scalar / vector	discrete-time system poles and residue vectors
d, $\boldsymbol{\Lambda}, \bar{\boldsymbol{B}}, \bar{\boldsymbol{C}}$	scalar / matrices	order and matrices of a state-space realization
$\mathbb{C}_T$	matrix $\in \mathbb{R}^{NC\times M}$	windowed controllability matrix (Remark 2)
B_z, B_W, B_a	scalar	bounds of Assumption 5
$\mathbb{E}$, Var, tr	operator	expectation, variance, trace

A.2 The Predictable Component Has a Convolution Structure

Proposition 1 (Duhamel–Volterra decomposition)**.** *Under Assumptions 1 and 3, the L_2-optimal predictor of the next-step field from $(\boldsymbol{S}, \mathcal{G})$ is*

$$\hat{\boldsymbol{u}}^*(\boldsymbol{S}) = \mathbb{E}[\boldsymbol{u}(T{+}1) \mid \boldsymbol{S}, \mathcal{G}] = \boldsymbol{A}^{T+1}\mathbb{E}[\boldsymbol{u}_0] + \sum_{m\ge 1} \sum_{1\le\tau_1,\dots,\tau_m\le T} \mathbf{G}_m\big[\boldsymbol{s}_{\tau_1}, \dots, \boldsymbol{s}_{\tau_m}\big], \tag{15}$$

where $\mathbf{G}_m$ is the order-m Volterra kernel (an m-linear map from $(\mathbb{R}^{D_s})^m$ to $\mathbb{R}^{NC}$), and $\mathbf{G}_1[\boldsymbol{s}_\tau] = \boldsymbol{A}^{T+1-\tau}\boldsymbol{B}\boldsymbol{s}_\tau$ is the causal Green kernel of the linearized system. In the linear case $\boldsymbol{n} \equiv 0$ only the $m = 1$ term survives.

Proof. Linear part. With $\boldsymbol{n} \equiv 0$, variation of constants applied to Eq. equation 14 over one step gives $\boldsymbol{u}(t{+}1) = \boldsymbol{A}\boldsymbol{u}(t) + \boldsymbol{B}\boldsymbol{s}_t$ in discrete time, hence by induction

$$\boldsymbol{u}(T{+}1) = \boldsymbol{A}^{T+1}\boldsymbol{u}_0 + \sum_{\tau=1}^{T} \boldsymbol{A}^{T+1-\tau}\boldsymbol{B}\boldsymbol{s}_\tau. \tag{16}$$

By Assumption 3 the initial-state-driven term is independent of $\boldsymbol{S}$, so taking conditional expectations eliminates its centered part and yields the $m = 1$ truncation of Eq. equation 15 plus the constant $\boldsymbol{A}^{T+1}\mathbb{E}[\boldsymbol{u}_0]$.

Nonlinear part. Write the solution of Eq. equation 14 via the Duhamel formula $\boldsymbol{u}(t) = \boldsymbol{A}^t\boldsymbol{u}_0 + \sum_{\tau\le t}\boldsymbol{A}^{t-\tau}\big[\boldsymbol{B}\boldsymbol{s}_\tau + \boldsymbol{n}(\boldsymbol{u}(\tau))\big]$ and expand $\boldsymbol{n}(\boldsymbol{u}) = \sum_{j\ge 2}\mathbb{N}_j[\boldsymbol{u},\dots,\boldsymbol{u}]$. Iterate: substituting the order-$(m{-}1)$ expansion of $\boldsymbol{u}$ into each $\mathbb{N}_j$ and collecting terms of total degree m in the observed

states defines $\mathbf{G}_m$ recursively, e.g. $\mathbf{G}_2[\boldsymbol{s}_{\tau_1}, \boldsymbol{s}_{\tau_2}] = \sum_t \boldsymbol{A}^{T+1-t}\,\mathbb{N}_2\big[\boldsymbol{A}^{t-\tau_1}\boldsymbol{B}\boldsymbol{s}_{\tau_1},\, \boldsymbol{A}^{t-\tau_2}\boldsymbol{B}\boldsymbol{s}_{\tau_2}\big]$. Each $\mathbf{G}_m$ is a finite sum of causal convolutions of lower-order kernels, hence measurable with respect to $\boldsymbol{S}$; convergence of the series follows from analyticity of $\boldsymbol{f}$ and Gronwall's inequality on the compact window. Since every term is $\sigma(\boldsymbol{S}, \mathcal{G})$-measurable and the residual is independent of $\boldsymbol{S}$ by Assumption 3, the conditional expectation selects exactly the right-hand side of Eq. equation 15. □

Remark 1 (What the no-field-history contract forfeits, and what it does not). Two statements must be separated. (i) *Within the sensor channel, nothing is forfeited.* Proposition 1 is exact over the admissible class: the L_2-optimal sensor-driven predictor depends on the input only through causal convolutions of $\boldsymbol{S}$, and no architectural modification can extract more from the same channel. The only field component outside the predictor's reach is the free response $\boldsymbol{A}^{T+1}(\boldsymbol{u}_0 - \mathbb{E}\boldsymbol{u}_0)$, which by Assumption 3 is independent of $\boldsymbol{S}$ and is therefore invisible to *every* predictor of the form $\hat{\boldsymbol{u}}(\boldsymbol{S}, \mathcal{G})$, regardless of capacity. (ii) *Across observation channels, the contract is not free.* Were field history available, a predictor could propagate the free response explicitly—in the linear case $\boldsymbol{u}(T{+}1) = \boldsymbol{A}\boldsymbol{u}(T) + \boldsymbol{B}\boldsymbol{s}_T$ would be computable up to the next-step forcing—and the floor of Proposition 2 would not apply. The exclusion of field history is thus a genuine information restriction, imposed not for convenience but by the deployment setting that motivates this work: operating edge equipment carries no in-domain field instrumentation at any time. The resulting error floor should be read as a quantitative, falsifiable price of the *observation channel* chosen by the hardware, rather than as a deficiency of any particular predictor; Proposition 2 makes this price exact, and Remark 2 shows it is structural whenever $M \ll NC$.

A.3 Observability and the Information Budget

Definition 1 (Predictable subspace)**.** The predictable (observable) subspace $\mathbb{U}_{\mathrm{obs}} \subseteq \mathbb{R}^{NC}$ is the closed linear span of all components of $\boldsymbol{u}(T{+}1)$ that are measurable with respect to $(\boldsymbol{S}, \mathcal{G})$; $\boldsymbol{P}$ and $\boldsymbol{P}_\perp = \boldsymbol{I} - \boldsymbol{P}$ are the orthogonal projectors onto $\mathbb{U}_{\mathrm{obs}}$ and its complement. This is the deterministic counterpart of the classical observable subspace of Kalman (1963).

Proposition 2 (Observability error floor)**.** *Under Assumptions 1–3, any predictor $\hat{\boldsymbol{u}}(\boldsymbol{S}, \mathcal{G})$ satisfies*

$$\inf_{\hat{\boldsymbol{u}}}\ \mathbb{E}\big\|\boldsymbol{u} - \hat{\boldsymbol{u}}\big\|_2^2\ \geq\ \mathbb{E}\big\|\boldsymbol{P}_\perp \boldsymbol{u}\big\|_2^2, \tag{17}$$

i.e. the optimal error is lower-bounded by the field energy inside the orthogonal complement of the predictable subspace.

Proof. Let $\boldsymbol{m}(\boldsymbol{S}) \triangleq \mathbb{E}[\boldsymbol{u} \mid \boldsymbol{S}, \mathcal{G}]$ be the conditional mean, which is the L_2-optimal predictor. For any $\hat{\boldsymbol{u}}(\boldsymbol{S}, \mathcal{G})$, expand

$$\mathbb{E}\|\boldsymbol{u} - \hat{\boldsymbol{u}}\|_2^2 = \mathbb{E}\|\boldsymbol{u} - \boldsymbol{m}\|_2^2 + \mathbb{E}\|\boldsymbol{m} - \hat{\boldsymbol{u}}\|_2^2 + 2\,\mathbb{E}\big\langle \boldsymbol{u} - \boldsymbol{m},\, \boldsymbol{m} - \hat{\boldsymbol{u}}\big\rangle. \tag{18}$$

The cross term vanishes: $\boldsymbol{m} - \hat{\boldsymbol{u}}$ is $\sigma(\boldsymbol{S}, \mathcal{G})$-measurable and $\mathbb{E}[\boldsymbol{u} - \boldsymbol{m} \mid \boldsymbol{S}, \mathcal{G}] = 0$, so $\mathbb{E}\langle \boldsymbol{u} - \boldsymbol{m}, \boldsymbol{m} - \hat{\boldsymbol{u}}\rangle = \mathbb{E}\langle \mathbb{E}[\boldsymbol{u} - \boldsymbol{m} \mid \boldsymbol{S}, \mathcal{G}], \boldsymbol{m} - \hat{\boldsymbol{u}}\rangle = 0$. Hence the infimum is attained at $\hat{\boldsymbol{u}} = \boldsymbol{m}$ with value $\mathbb{E}\|\boldsymbol{u} - \boldsymbol{m}\|_2^2$. By Assumption 3, $\mathbb{E}[\boldsymbol{P}_\perp \boldsymbol{u} \mid \boldsymbol{S}, \mathcal{G}] = \mathbb{E}[\boldsymbol{P}_\perp \boldsymbol{u}] = 0$, so $\boldsymbol{P}_\perp \boldsymbol{m} = 0$ and

$$\mathbb{E}\|\boldsymbol{u} - \boldsymbol{m}\|_2^2 = \mathbb{E}\|\boldsymbol{P}(\boldsymbol{u} - \boldsymbol{m})\|_2^2 + \mathbb{E}\|\boldsymbol{P}_\perp \boldsymbol{u}\|_2^2\ \geq\ \mathbb{E}\|\boldsymbol{P}_\perp \boldsymbol{u}\|_2^2, \tag{19}$$

where the first equality is the Pythagorean theorem for orthogonal projectors. □

Remark 2 (Budget-induced dimension bound: a falsifiable saturation prediction). The information budget constrains the predictable subspace through the *rank of the observation map*, not merely through the scalar count M. For the linearized dynamics ($\boldsymbol{n} \equiv 0$), the forced component of Eq. equation 16 is a linear image of the window: it ranges over the column space of the windowed controllability matrix

$$\mathbb{C}_T \triangleq \big[\, \boldsymbol{A}^T\boldsymbol{B},\ \boldsymbol{A}^{T-1}\boldsymbol{B},\ \ldots,\ \boldsymbol{A}\boldsymbol{B}\,\big] \in \mathbb{R}^{NC\times M}, \qquad \dim \mathbb{U}_{\mathrm{obs}}\ \leq\ \operatorname{rank}\mathbb{C}_T\ \leq\ \min(M,\, NC). \tag{20}$$

For the nonlinear dynamics, the order-m Volterra term $\boldsymbol{s} \mapsto \mathbf{G}_m[\boldsymbol{s}, \ldots, \boldsymbol{s}]$ is a polynomial map of degree m, so its range lies in the column space of the symmetrized unfolding of $\mathbf{G}_m$, of dimension at most $\binom{M+m-1}{m}$; truncating the series at order m_0 therefore gives

$$\dim \mathbb{U}_{\mathrm{obs}}\ \leq\ \sum_{m=1}^{m_0} \binom{M+m-1}{m} = \binom{M+m_0}{m_0} - 1. \tag{21}$$

On our benchmarks $M \in \{300,\ 850\}$ while $NC \approx 1.4\times10^4$–1.4×10^5, so the predictable subspace remains a thin slice of field space for any moderate m_0. Consequently, whenever the field energy spectrum decays slowly, $\mathbb{E}\|\boldsymbol{P}_\perp \boldsymbol{u}\|_2^2$ is a non-negligible fraction of the total energy, and Proposition 2 predicts that *every* predictor constrained by the same budget saturates at a common error level regardless of architecture. This saturation is directly testable: lengthening the observation window enlarges $\mathbb{C}_T$ and lowers the floor, which is precisely the behavior observed in the window-length sweep of our structural ablations, where the error approaches, but never crosses, the Wiener-estimated floor.

A.4 Fixed Bases Are Structurally Suboptimal

Under Assumption 4, the L_2-optimal predictor conditioned on g is a conditional Karhunen–Loève expansion: for each g, take the top-r eigenvectors of $\boldsymbol{C}_g$ as bases and conditional projections as coefficients. A *fixed-basis* predictor instead uses one frozen rank-r basis $\boldsymbol{V}$ for all conditions. The next proposition quantifies the gap.

Proposition 3 (Irreducible error of fixed bases)**.** *Let $\boldsymbol{C} \triangleq \mathbb{E}[\boldsymbol{u}\boldsymbol{u}^\top] = \mathbb{E}_g[\boldsymbol{C}_g] + \boldsymbol{M}$ be the mixture second-moment matrix (law of total covariance, with $\boldsymbol{M} = \mathrm{Var}_g[\boldsymbol{\mu}_g]$). Then*

$$\underbrace{\mathrm{tr}(\boldsymbol{C}) - \sum_{i\le r}\lambda_i(\boldsymbol{C})}_{\text{best fixed rank-}r\text{ error}} \;\ge\; \underbrace{\mathbb{E}_g\Big[\,\mathrm{tr}(\boldsymbol{C}_g) - \sum_{i\le r}\lambda_i(\boldsymbol{C}_g)\Big]}_{\text{conditional rank-}r\text{ error}} + \Big(1 - \frac{r}{r_{\mathrm{eff}}}\Big)\,\mathrm{tr}(\boldsymbol{M})\,\big(1 - o(1)\big). \tag{22}$$

The last term is independent of the basis and cannot be removed by increasing r while $r \ll r_{\mathrm{eff}}$; a conditional basis choosing, for each instance, its own Eckart–Young-optimal subspace annihilates it.

Proof. For any fixed basis $\boldsymbol{V}$ with orthonormal columns, the expected projection error is $\mathbb{E}\|\boldsymbol{u} - \boldsymbol{V}\boldsymbol{V}^\top\boldsymbol{u}\|_2^2 = \mathrm{tr}(\boldsymbol{C}) - \mathrm{tr}(\boldsymbol{V}^\top\boldsymbol{C}\boldsymbol{V})$, minimized when $\mathrm{tr}(\boldsymbol{V}^\top\boldsymbol{C}\boldsymbol{V})$ is maximal. By the Ky Fan maximum principle, $\max_{\boldsymbol{V}:\,\boldsymbol{V}^\top\boldsymbol{V}=\boldsymbol{I}_r}\mathrm{tr}(\boldsymbol{V}^\top\boldsymbol{C}\boldsymbol{V}) = \sum_{i\le r}\lambda_i(\boldsymbol{C})$, which proves the left-hand side is the best fixed-basis error. Now expand, for any such $\boldsymbol{V}$,

$$\mathrm{tr}(\boldsymbol{V}^\top\boldsymbol{C}\boldsymbol{V}) = \mathbb{E}_g\big[\,\mathrm{tr}(\boldsymbol{V}^\top\boldsymbol{C}_g\boldsymbol{V})\big] + \mathrm{tr}(\boldsymbol{V}^\top\boldsymbol{M}\boldsymbol{V}). \tag{23}$$

For the first term, Ky Fan applied pointwise in g gives $\mathrm{tr}(\boldsymbol{V}^\top\boldsymbol{C}_g\boldsymbol{V}) \le \sum_{i\le r}\lambda_i(\boldsymbol{C}_g)$ for every $\boldsymbol{V}$ and every g, hence $\mathbb{E}_g[\mathrm{tr}(\boldsymbol{V}^\top\boldsymbol{C}_g\boldsymbol{V})] \le \mathbb{E}_g\big[\sum_{i\le r}\lambda_i(\boldsymbol{C}_g)\big]$. For the second term, $\mathrm{tr}(\boldsymbol{V}^\top\boldsymbol{M}\boldsymbol{V}) \le \sum_{i\le r}\lambda_i(\boldsymbol{M}) \le r\,\lambda_1(\boldsymbol{M}) \le \big(r/r_{\mathrm{eff}}\big)\,\mathrm{tr}(\boldsymbol{M})\,(1 + o(1))$ by the flat-spectrum condition of Assumption 4. Substituting both bounds into Eq. equation 23 and taking the maximum over $\boldsymbol{V}$,

$$\sum_{i\le r}\lambda_i(\boldsymbol{C}) \le \mathbb{E}_g\Big[\sum_{i\le r}\lambda_i(\boldsymbol{C}_g)\Big] + \frac{r}{r_{\mathrm{eff}}}\,\mathrm{tr}(\boldsymbol{M})\,(1 + o(1)). \tag{24}$$

Finally, $\mathrm{tr}(\boldsymbol{C}) = \mathbb{E}_g[\mathrm{tr}(\boldsymbol{C}_g)] + \mathrm{tr}(\boldsymbol{M})$ by the law of total covariance; subtracting the inequality above and rearranging yields Eq. equation 22. □

Remark 3 (Why marginal energy capture is the wrong diagnostic). For any frozen orthonormal basis $\boldsymbol{V}$ and coefficient estimate $\boldsymbol{a}(\boldsymbol{S})$, the prediction error decomposes orthogonally as

$$\mathbb{E}\big\|\boldsymbol{u} - \boldsymbol{V}\boldsymbol{a}\big\|_2^2 = \underbrace{\mathbb{E}\big\|(\boldsymbol{I} - \boldsymbol{V}\boldsymbol{V}^\top)\boldsymbol{u}\big\|_2^2}_{\text{marginal projection deficit}} + \underbrace{\mathbb{E}\big\|\boldsymbol{V}^\top\boldsymbol{u} - \boldsymbol{a}(\boldsymbol{S})\big\|_2^2}_{\text{coefficient regression error}}, \tag{25}$$

since $(\boldsymbol{I} - \boldsymbol{V}\boldsymbol{V}^\top)\boldsymbol{u} \perp \boldsymbol{V}(\boldsymbol{V}^\top\boldsymbol{u} - \boldsymbol{a})$. On the elastic benchmark the first term is small: a rank-2 POD basis captures 99.95% of the training energy, a marginal projection deficit of 5×10^{-4} in relative terms, while a tuned frozen-basis predictor (SHRED) attains 0.039 test NMSE—nearly two orders of magnitude above the marginal floor. Two properties of Eq. equation 25 resolve the apparent tension. First, the decomposition is evaluated on the *mixture* distribution, whereas prediction is *conditional*: the directions that minimize the marginal deficit are averaged across states, and Proposition 3 quantifies exactly how much a single basis loses by this averaging, namely the across-state variation term $(1 - r/r_{\mathrm{eff}})\,\mathrm{tr}(\boldsymbol{M})$, which a per-instance basis removes. Second, the 99.95% figure is computed in physical energy units and is dominated by the highest-variance channels and directions, whereas the reported NMSE is computed in the standardized field space in which every channel contributes

equally; a frozen basis can therefore saturate marginal energy capture while remaining misaligned with the channel-balanced, per-instance directions that the evaluation metric rewards. Consistently, enlarging the rank of a frozen basis does not help (the added marginal directions are not the missing conditional ones), while the state-conditioning ablation isolates the directional term directly: replacing the generated basis by an equal-rank frozen POD basis degrades test NMSE from 0.0249 to 0.0612 on the same benchmark, a gap far exceeding the marginal projection deficit and attributable precisely to the missing per-instance alignment.

A.5 The Conditional Spectral Hypothesis Class

Definition 2. The fixed-basis class is $\mathbb{H}_{\mathrm{POD}} = \{\boldsymbol{u} \mapsto \boldsymbol{V}\boldsymbol{a}_\psi(\boldsymbol{S})\}$ with frozen $\boldsymbol{V}$; the conditional spectral class is $\mathbb{H}_{\mathrm{SONAR}} = \{(\boldsymbol{S}, \mathcal{G}) \mapsto \boldsymbol{\Phi}_\theta(\mathcal{G}; \boldsymbol{z}_s)\, \boldsymbol{a}_\psi(\boldsymbol{z}_s)\}$, where $\boldsymbol{\Phi}_\theta \in \mathbb{R}^{NC\times R}$ is produced per sample by message passing on $\mathcal{G}$ conditioned on the sensor embedding $\boldsymbol{z}_s$, and $\boldsymbol{a}_\psi \in \mathbb{R}^R$ is produced from $\boldsymbol{z}_s$.

Proposition 4 (Strict nesting and monotone warm-start improvement)**.** *(i) $\mathbb{H}_{\mathrm{POD}} \subsetneq \mathbb{H}_{\mathrm{SONAR}}$. (ii) Initializing $\boldsymbol{\Phi}_\theta$ at a frozen POD basis and training by full-batch gradient descent with step sizes satisfying the descent lemma ($\eta \le 1/L$, where L is the smoothness constant of $\mathbb{L}$) yields training error no larger than at initialization; under mini-batch stochastic gradients the same statement holds in expectation over the sampling noise. (iii) The expected gain of the conditional class over its POD sub-model is the variance term of Proposition 3, of order $\big(1 - r/r_{\mathrm{eff}}\big)\,\mathrm{tr}(\boldsymbol{M})$.*

Proof. (i) Taking $\boldsymbol{\Phi}_\theta(\cdot) \equiv \boldsymbol{V}$ constant realizes any element of $\mathbb{H}_{\mathrm{POD}}$, giving inclusion. Strictness follows from a two-condition example: let $\boldsymbol{C}_{g_1} = \sigma^2 \boldsymbol{e}_1\boldsymbol{e}_1^\top$ and $\boldsymbol{C}_{g_2} = \sigma^2 \boldsymbol{e}_2\boldsymbol{e}_2^\top$ with $\boldsymbol{e}_1 \perp \boldsymbol{e}_2$, equal mixing weights, and $r = 1$. A conditional basis achieves zero error, while any fixed rank-1 basis $\boldsymbol{v}$ incurs $\frac{\sigma^2}{2}\big(1 - \langle \boldsymbol{v}, \boldsymbol{e}_1\rangle^2\big) + \frac{\sigma^2}{2}\big(1 - \langle \boldsymbol{v}, \boldsymbol{e}_2\rangle^2\big) = \frac{\sigma^2}{2} > 0$. (ii) With step sizes satisfying the descent lemma, each gradient update does not increase the training objective $\mathbb{L}$: $\mathbb{L}(\theta_{k+1}) \le \mathbb{L}(\theta_k)$; chaining from the POD initialization θ_0 gives $\mathbb{L}(\theta_K) \le \mathbb{L}(\theta_0)$. For stochastic gradients, $\mathbb{E}[\mathbb{L}(\theta_{k+1}) \mid \theta_k] \le \mathbb{L}(\theta_k) + \frac{L\eta^2}{2}\mathbb{E}\|\boldsymbol{g}_k\|^2$ with unbiased gradient estimates $\boldsymbol{g}_k$, so the expected objective is non-increasing up to the standard $O(L\eta^2)$ noise term, which vanishes as $\eta \to 0$. (iii) Subtract the conditional error from the fixed-basis error in Eq. equation 22. □

Proposition 5 (Whitening identifies the coefficients)**.** *For the bilinear model $\hat{\boldsymbol{u}} = \boldsymbol{\Phi}\boldsymbol{a}$: (i) without constraints, $(\boldsymbol{\Phi}\boldsymbol{W},\ \boldsymbol{W}^{-1}\boldsymbol{a})$ gives identical predictions for every $\boldsymbol{W} \in GL(R)$, so $\boldsymbol{a}$ is not identifiable; (ii) imposing the Gram-whitening constraint $\boldsymbol{\Phi}^\top\boldsymbol{\Phi}/(NC) = \boldsymbol{I}_R$ restricts the gauge group to the orthogonal group $O(R)$, and coefficients become identifiable up to sign and permutation; (iii) for the least-squares loss $\mathbb{L}(\boldsymbol{a}) = \|\boldsymbol{u} - \boldsymbol{\Phi}\boldsymbol{a}\|_F^2$, exact whitening makes the coefficient-side curvature perfectly isotropic, $\nabla^2_{\boldsymbol{a}}\mathbb{L} = 2\,\boldsymbol{\Phi}^\top\boldsymbol{\Phi} = 2NC\,\boldsymbol{I}_R$; under the jittered whitening $\boldsymbol{\Phi} \leftarrow \boldsymbol{\Phi}\big(\boldsymbol{\Phi}^\top\boldsymbol{\Phi}/(NC) + \epsilon\boldsymbol{I}\big)^{-1/2}$ used in practice, the Hessian equals $2NC\,\boldsymbol{U}\,\mathrm{diag}\big(\lambda_i/(\lambda_i+\epsilon)\big)\,\boldsymbol{U}^\top$, where $\boldsymbol{U}\mathrm{diag}(\lambda_i)\boldsymbol{U}^\top$ is the eigendecomposition of the pre-whitening Gram matrix $\boldsymbol{\Phi}^\top\boldsymbol{\Phi}/(NC)$, so its condition number is $\kappa = \frac{\lambda_1}{\lambda_R}\cdot\frac{\lambda_R+\epsilon}{\lambda_1+\epsilon} = 1 + O(\epsilon/\lambda_R)$, recovering exact isotropy as $\epsilon \to 0$.*

Proof. (i) $\boldsymbol{\Phi}\boldsymbol{W}\boldsymbol{W}^{-1}\boldsymbol{a} = \boldsymbol{\Phi}\boldsymbol{a}$ for any invertible $\boldsymbol{W}$; the continuous symmetry group $GL(R)$ acts trivially on predictions but transitively on $(\boldsymbol{\Phi}, \boldsymbol{a})$, so no estimator can separate them. (ii) The whitening constraint defines a Stiefel manifold in $\boldsymbol{\Phi}$-space. A gauge map $\boldsymbol{W}$ preserves the constraint iff $(\boldsymbol{\Phi}\boldsymbol{W})^\top(\boldsymbol{\Phi}\boldsymbol{W})/(NC) = \boldsymbol{W}^\top\boldsymbol{W} = \boldsymbol{I}_R$, i.e. $\boldsymbol{W} \in O(R)$; if in addition the coefficients are ordered by magnitude, the residual freedom is a signed permutation, a finite group, which is the classical notion of identifiability up to label symmetry. Geometrically the parameterization becomes a fiber bundle with Stiefel base and compact $O(R)$ fiber, quotienting out the noncompact part of $GL(R)$. (iii) $\nabla_{\boldsymbol{a}}\mathbb{L} = 2\boldsymbol{\Phi}^\top(\boldsymbol{\Phi}\boldsymbol{a} - \boldsymbol{u})$ and $\nabla^2_{\boldsymbol{a}}\mathbb{L} = 2\boldsymbol{\Phi}^\top\boldsymbol{\Phi}$; under exact whitening $\boldsymbol{\Phi}^\top\boldsymbol{\Phi} = NC\,\boldsymbol{I}_R$, so the Hessian has condition number 1 and every coefficient receives gradients of equal scale. With jitter, writing $\boldsymbol{G} \triangleq \boldsymbol{\Phi}^\top\boldsymbol{\Phi}/(NC) = \boldsymbol{U}\mathrm{diag}(\lambda_i)\boldsymbol{U}^\top$, the whitened basis $\boldsymbol{\Phi}_\epsilon = \boldsymbol{\Phi}(\boldsymbol{G} + \epsilon\boldsymbol{I})^{-1/2}$ satisfies $\boldsymbol{\Phi}_\epsilon^\top\boldsymbol{\Phi}_\epsilon = NC\,\boldsymbol{U}\,\mathrm{diag}\big(\lambda_i/(\lambda_i+\epsilon)\big)\,\boldsymbol{U}^\top$, whose eigenvalues lie in $(0, NC]$ and whose condition number is $\frac{\lambda_1/(\lambda_1+\epsilon)}{\lambda_R/(\lambda_R+\epsilon)} = \frac{\lambda_1(\lambda_R+\epsilon)}{\lambda_R(\lambda_1+\epsilon)} = 1 + O(\epsilon/\lambda_R)$. □

Proposition 6 (Universal approximation is preserved)**.** *If the target operator $(\boldsymbol{S}, \mathcal{G}) \mapsto \boldsymbol{u}$ is continuous on a compact input set, $\mathbb{H}_{\mathrm{SONAR}}$ approximates it uniformly to arbitrary accuracy for sufficiently large R and message-passing depth.*

Proof sketch. Branch–trunk architectures $\sum_{r=1}^{R} b_r(\boldsymbol{S})\, t_r(\cdot)$ are universal for continuous operators (Lu et al., 2021). It remains to note that replacing the pointwise trunk t_r by a message-passing basis generator does not shrink the class: with injective coordinate node features and sufficient depth, message-passing networks approximate any continuous permutation-equivariant function on attributed graphs up to the resolution of the Weisfeiler–Leman test (Xu et al., 2019), which unique coordinates refine to the level of individual nodes; hence the basis family $\{\boldsymbol{\Phi}_\theta\}$ is dense in the space of continuous maps from $\mathcal{G}$ to $\mathbb{R}^{NC\times R}$. Products of dense families with dense coefficient maps are dense under the supremum norm by a Stone–Weierstrass argument, completing the proof. □

A.6 Supervision Structure and Sample Efficiency

Proposition 7 (Asymmetric supervision)**.** *Each training sample provides NC supervised scalars. In $\mathbb{H}_{\mathrm{SONAR}}$ the geometry-to-basis map is supervised at every node for every basis field, an effective supervision volume of order NR per sample, while the sensor-to-coefficient map produces only $R \ll NC$ numbers. Moreover, under Assumption 5, the coefficient layer $\boldsymbol{a} = \boldsymbol{W}\boldsymbol{z}_s$ admits a norm-based generalization bound of order $\tilde{\mathbb{O}}\big((B_z B_W + B_a)\, B_z B_W/\sqrt{n}\big)$; relative to a linear decoder emitting all NC outputs under a matched per-row norm budget, the bound is smaller by a factor of order $\sqrt{R/(NC)}$.*

Proof sketch. The supervision count is direct: the basis fields enter the loss at every node, whereas the coefficients enter only through R inner products per sample. For the bound, consider the linear class $\mathbb{F} = \{\boldsymbol{z} \mapsto \boldsymbol{W}\boldsymbol{z} : \|\boldsymbol{W}\|_F \le B_W\}$ with $\|\boldsymbol{z}\|_2 \le B_z$. Its empirical Rademacher complexity satisfies

$$\hat{\mathbb{R}}_n(\mathbb{F}) = \frac{1}{n}\,\mathbb{E}_{\boldsymbol{\epsilon}} \sup_{\|\boldsymbol{W}\|_F \le B_W} \Big\| \sum_{i=1}^{n} \varepsilon_i\, \boldsymbol{W}\boldsymbol{z}_i \Big\|_2 \le \frac{B_W}{n}\, \mathbb{E}_{\boldsymbol{\epsilon}} \Big\| \sum_{i=1}^{n} \varepsilon_i \boldsymbol{z}_i \Big\|_2 \le \frac{B_W B_z}{\sqrt{n}}, \tag{26}$$

where the first inequality uses $\|\boldsymbol{W}\boldsymbol{z}\|_2 \le \|\boldsymbol{W}\|_F \|\boldsymbol{z}\|_2$ and the second uses $\mathbb{E}\|\sum_i \varepsilon_i \boldsymbol{z}_i\|_2 \le \big(\sum_i \|\boldsymbol{z}_i\|_2^2\big)^{1/2} \le \sqrt{n}\, B_z$ by orthogonality of Rademacher signs. The squared loss $\ell(\boldsymbol{a}, \hat{\boldsymbol{a}}) = \|\boldsymbol{a} - \hat{\boldsymbol{a}}\|_2^2$ on the bounded domain has $\nabla_{\hat{\boldsymbol{a}}}\ell = 2(\hat{\boldsymbol{a}} - \boldsymbol{a})$, hence is Lipschitz in the Euclidean norm with constant $2(B_z B_W + B_a)$, *independent of the output dimension*; Talagrand's contraction lemma and symmetrization then give the stated rate. To compare output dimensions on equal footing, match the per-row norm budget: if every output row of $\boldsymbol{W}$ has norm at most b, then $B_W = b\sqrt{R}$ for the coefficient head and $B_W = b\sqrt{NC}$ for a full-field decoder, and Eq. equation 26 scales as $\sqrt{R}$ versus $\sqrt{NC}$, a factor of order $\sqrt{R/(NC)}$. □

Remark 4. Proposition 7 prescribes an asymmetric architecture: depth is safe on the geometry side, where supervision is dense, and wasteful on the sensor side, where the output is R-dimensional. This is consistent with the model tolerating a 15-step basis generator at $n = 375$ training samples without the overfitting observed in attention-based global models, and with message-passing depth being the first-order accuracy knob in our ablations.

A.7 Realization Theory of the Green Kernel

Proposition 8 (State-space realization and FIR truncation)**.** *Let the discrete-time linear Green kernel be $g_k = \bar{\boldsymbol{C}}\boldsymbol{\Lambda}^{k-1}\bar{\boldsymbol{B}}$, $k \ge 1$, with poles $\{\lambda_p\}_{p=1}^{d}$ and residue vectors $\{\boldsymbol{c}_p\}$ (complex poles occurring in conjugate pairs). (i) There exists a minimal state-space realization*

$$\boldsymbol{h}_{t+1} = \boldsymbol{\Lambda}\boldsymbol{h}_t + \bar{\boldsymbol{B}}\boldsymbol{s}_t, \qquad \boldsymbol{a}_t = \bar{\boldsymbol{C}}\boldsymbol{h}_t, \qquad \boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_d), \tag{27}$$

of order d equal to the rank of the Hankel matrix of $\{g_k\}$, computing the convolution of Proposition 1 over arbitrarily long windows in $O(Td)$ time and $O(d)$ memory (Ho & Kálmán, 1966). (ii) Truncating the kernel at length K, i.e. keeping $g_1, \ldots, g_K$, incurs tail energy

$$\|g - g_{\le K}\|_2^2 = \sum_{k>K} \Big\| \sum_{p=1}^{d} \boldsymbol{c}_p\, \lambda_p^{k-1} \Big\|_2^2 = \sum_{p,q=1}^{d} \boldsymbol{c}_p^{*}\boldsymbol{c}_q\, \frac{(\lambda_p^{*}\lambda_q)^K}{1 - \lambda_p^{*}\lambda_q} \;\le\; d \sum_{p=1}^{d} \frac{\|\boldsymbol{c}_p\|_2^2\, |\lambda_p|^{2K}}{1 - |\lambda_p|^2}, \tag{28}$$

where the equality is exact (the double sum is real because the summand is Hermitian in (p, q)) and the inequality is Cauchy–Schwarz over the d modes. When the residue vectors are mutually orthogonal, the cross terms vanish and Eq. equation 28 reduces to the diagonal form $\sum_p \|\boldsymbol{c}_p\|_2^2 \, |\lambda_p|^{2K}/(1-|\lambda_p|^2)$.

Proof. (i) Form the block Hankel matrix $\boldsymbol{H} = (g_{i+j-1})_{i,j}$ and its shift $\boldsymbol{H}' = (g_{i+j})_{i,j}$. Factor $\boldsymbol{H} = \mathbb{O}\,\mathbb{C}$ with $d = \operatorname{rank} \boldsymbol{H}$ (observability and controllability factors); the shift structure of the Hankel matrix gives $\boldsymbol{H}' = \mathbb{O}\,\boldsymbol{\Lambda}\,\mathbb{C}$, so solving $\boldsymbol{\Lambda} = \mathbb{O}^\dagger \boldsymbol{H}' \mathbb{C}^\dagger$ recovers the transition matrix, while $\bar{\boldsymbol{B}}$ and $\bar{\boldsymbol{C}}$ are read from the first block column of $\mathbb{C}$ and the first block row of $\mathbb{O}$, respectively. This is the Ho–Kalman construction (Ho & Kálmán, 1966); the recursion equation 27 then evaluates $\sum_{\tau \le t} g_{t-\tau} \boldsymbol{s}_\tau$ online with one state update per step, using $O(d)$ memory. (ii) By diagonalizability, $g_k = \sum_p \boldsymbol{c}_p \lambda_p^{k-1}$. Expanding the squared norm, summing the geometric series over $k > K$—valid since $|\lambda_p^* \lambda_q| \le (1-\varepsilon_0)^2 < 1$ by Assumption 2—gives the middle expression of Eq. equation 28: $\sum_{k>K} (\lambda_p^* \lambda_q)^{k-1} = (\lambda_p^* \lambda_q)^K/(1-\lambda_p^* \lambda_q)$. The inequality follows from $\big\| \sum_p \boldsymbol{c}_p \lambda_p^{k-1} \big\|_2^2 \le d \sum_p \|\boldsymbol{c}_p\|_2^2 |\lambda_p|^{2(k-1)}$ (Cauchy–Schwarz) summed over k. If $\boldsymbol{c}_p^* \boldsymbol{c}_q = 0$ for $p \ne q$, only the diagonal terms survive. □

Remark 5. If a mode is weakly damped, $|\lambda_p| = 1-\epsilon$, its diagonal tail contribution is $\|\boldsymbol{c}_p\|_2^2 (1-\epsilon)^{2K}/\big(1-(1-\epsilon)^2\big) = \|\boldsymbol{c}_p\|_2^2 (1-\epsilon)^{2K}/\big(2\epsilon-\epsilon^2\big) = \Theta(1/\epsilon)$, so any short finite-impulse-response realization is structurally inadequate for elastic or diffusive systems, whose poles cluster near the unit circle. A gated recurrent update $\boldsymbol{h}_t = (1-\boldsymbol{\gamma}_t) \odot \boldsymbol{h}_{t-1} + \boldsymbol{\gamma}_t \odot \tilde{\boldsymbol{h}}_t$ with slowly varying gates is a nonlinear infinite-impulse-response filter with time-varying effective poles and therefore lies in the function class of Proposition 8(i), which justifies a GRU as the temporal encoder and explains why the choice among gated recurrent cells is a second-order effect in our experiments.

B Details of Datasets

We construct three synthetic datasets, ElasticBeam-Syn, HeatPlate-Syn and Go1Leg-Dyn, to instantiate the task of predicting the next full field snapshot from system-level observed state series on unstructured meshes. All three datasets follow one design protocol. Each sample consists of a multichannel sensor series of 50 time steps sampled at 100 Hz, an unstructured tetrahedral mesh, and the target physical field evaluated at the first step after the observation window. The sensed quantities are system level signals whose physical meaning differs from the target field, and no interior field measurement is ever exposed, which mirrors the virtual sensing setting in digital twin applications that motivates this work. Within each dataset the geometry, the material and the mesh are fixed, so the only source of variation across the 500 samples is the observed state history together with injected sensor noise and missing values. Each dataset is split into 375 training, 62 validation and 63 test samples with a fixed seed of 42, and all normalization statistics are computed on the training split only.

B.1 ElasticBeam-Syn

ElasticBeam-Syn models a three dimensional cantilever beam of length 3.0 m, width 0.5 m and height 0.2 m with a circular through hole of diameter 0.1 m centered 1.0 m away from the clamped end, discretized by Gmsh into an unstructured tetrahedral mesh of 2,388 nodes and 9,031 elements shared by all samples. The material is isotropic linear elastic with Young's modulus $E = 100$ GPa, Poisson ratio $\nu = 0.3$ and density $\rho = 7850\ \mathrm{kg\,m^{-3}}$. Each sample solves the linear elasticity boundary value problem

$$\nabla \cdot \boldsymbol{\sigma} = \mathbf{0}, \qquad \boldsymbol{\sigma} = \lambda \operatorname{tr}(\boldsymbol{\varepsilon})\, \mathbf{I} + 2\mu\, \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} = \tfrac{1}{2}\big(\nabla \mathbf{u} + \nabla \mathbf{u}^\top\big), \tag{29}$$

where $\mathbf{u}$ denotes the displacement field and λ and μ denote the Lamé constants derived from E and ν. The end at $x = 0$ is clamped with $\mathbf{u} = \mathbf{0}$, and the tip face at $x = L$ is driven by a traction whose resultant follows a damped oscillatory history

$$F(t) = F_0 \sin(2\pi f t + \phi)\, e^{-\lambda_d t} + \eta(t), \tag{30}$$

where $F_0 = 10^5$ N is the amplitude, $f = 1.0$ Hz is the frequency, $\lambda_d = 0.2$ is the decay rate and $\eta(t)$ is a sample wise random perturbation. The input series records the three displacement components

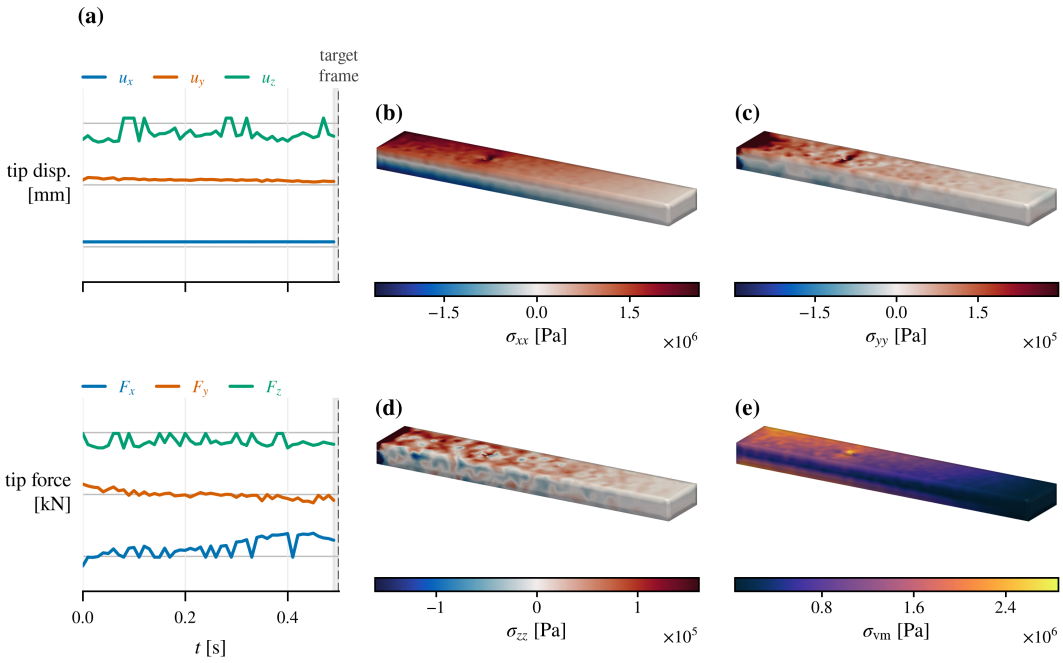

Figure 6: ElasticBeam-Syn. Panel a shows the 50 step input window of tip displacement in mm and tip force in kN, with the dashed line marking the target frame one step after the window. Panels b to e show the target frame stress field on the shared mesh, including the normal components σ_{xx}, σ_{yy} and σ_{zz} in Pa with symmetric color scales, and the von Mises stress σ_{vm} in Pa. The displayed sample is the median field energy sample of the test split, and color limits follow the 2% to 98% percentile range over the test split.

and the three force components measured at the center of the tip face, forming a 6 channel signal that parameterizes the loading state of the structure. The target field contains the six independent stress components $\sigma_{xx}, \sigma_{yy}, \sigma_{zz}, \sigma_{xy}, \sigma_{xz}, \sigma_{yz}$ at every mesh node one step after the window, computed with FEniCSx and projected onto the nodal space. The von Mises stress reaches 5.8 MPa over the dataset and concentrates around the hole and the clamped end, which makes this dataset a precise testbed for peak error behavior. Figure 6 illustrates one representative test sample.

B.2 HeatPlate-Syn

HeatPlate-Syn models transient heat conduction in a thin aluminum plate of length 2.0 m, width 1.0 m and thickness 0.03 m, discretized into 2,075 nodes and 5,925 tetrahedra shared by all samples. Two Gaussian heat sources of radii 0.1 m and 0.08 m are fixed in the mid plane of the plate at 0.8 m and 1.5 m along the length direction. The thermal conductivity is $k = 200\ \mathrm{W\,m^{-1}\,K^{-1}}$, the density is $\rho = 2700\ \mathrm{kg\,m^{-3}}$ and the specific heat is $c_p = 900\ \mathrm{J\,kg^{-1}\,K^{-1}}$. The temperature field T is governed by

$$\rho\, c_p\, \frac{\partial T}{\partial t} = \nabla\cdot(k\,\nabla T) + Q(\mathbf{x}, t), \tag{31}$$

with convective Robin conditions $-k\,\partial T/\partial n = h\,(T - T_{\mathrm{env}})$ on all outer surfaces, where $h = 50\ \mathrm{W\,m^{-2}\,K^{-1}}$ and the environment temperature T_{env} oscillates around a baseline of 350 K. The source term $Q(\mathbf{x}, t)$ superposes the two Gaussian profiles and is normalized so that the total injected power equals the sensor value $Q_{\mathrm{total}}(t)$ at every step. Each sample draws the boundary temperature amplitude between 10 K and 100 K, the total source power between 10^3 W and 5×10^4 W, and the fluctuation frequency between 0.05 Hz and 0.3 Hz. The system is integrated with the backward Euler scheme over 50 steps. The input series contains 6 channels, namely the normal heat flux at the midpoints of three lateral boundaries, the top and bottom boundary temperatures, and the total source power. The target field is the temperature gradient ∇T at the final step, which corresponds to

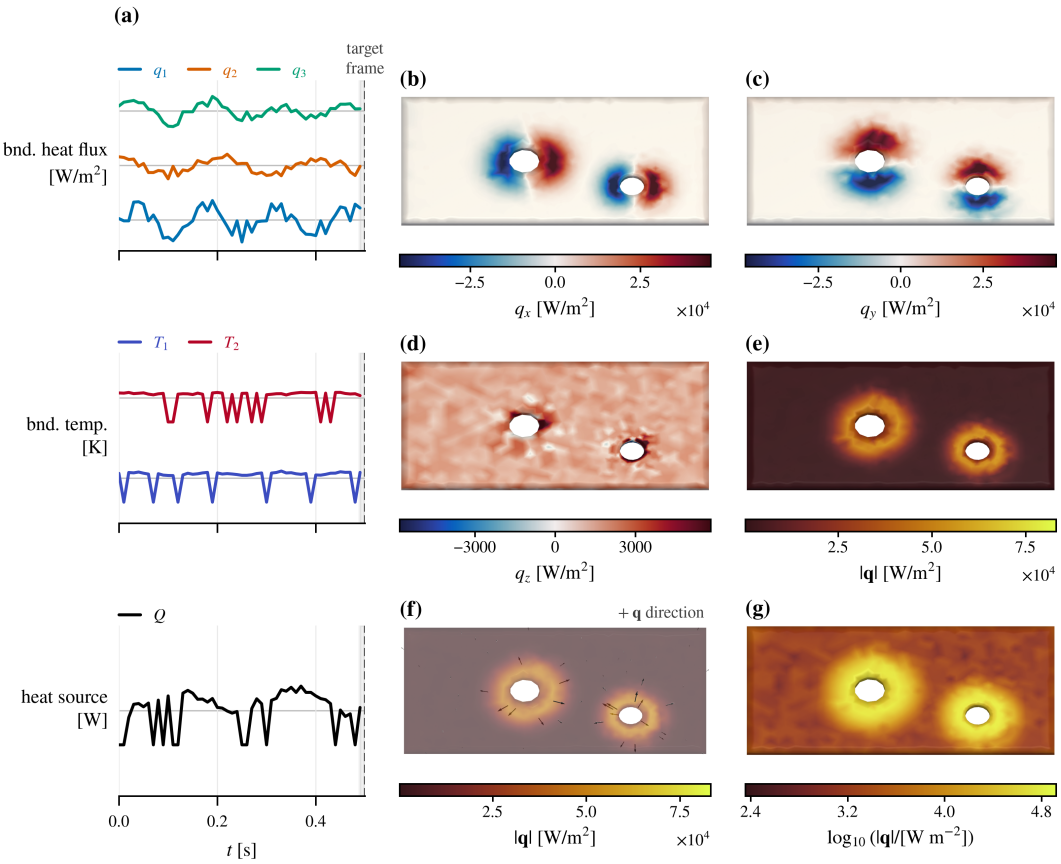

Figure 7: HeatPlate-Syn. Panel a shows the 50 step input window of boundary heat flux in $\mathrm{W\,m^{-2}}$, boundary temperature in K, and total source power in W. Panels b to d show the target frame heat flux components q_x, q_y and q_z in $\mathrm{W\,m^{-2}}$ with symmetric color scales, panel e shows the flux magnitude $|\mathbf{q}|$, panel f overlays subsampled direction glyphs on the magnitude field, and panel g shows the magnitude in $\log_{10}$ scale to expose the far field decay. The displayed sample is the median field energy sample of the test split.

the heat flux $\mathbf{q} = -k\,\nabla T$ through Fourier's law, with the flux magnitude reaching 1.4×10^5 $\mathrm{W\,m^{-2}}$ over the dataset. Figure 7 illustrates one representative test sample.

B.3 Go1Leg-Dyn

Go1Leg-Dyn is built from the front right leg of the Unitree Go1 quadruped, whose geometry derives from the publicly available simulation assets of MuJoCo Menagerie. The leg assembles the hip, thigh and calf components into one unstructured tetrahedral mesh of 22,515 nodes and 93,983 elements shared by all samples, with aluminum alloy material of $E = 70$ GPa, $\nu = 0.33$ and $\rho = 2700$ $\mathrm{kg\,m^{-3}}$, and Rayleigh damping at a 2% modal ratio. Data generation couples two simulators. MuJoCo integrates the multibody dynamics

$$M(\mathbf{q}_j)\,\ddot{\mathbf{q}}_j + \mathbf{c}(\mathbf{q}_j, \dot{\mathbf{q}}_j) + \mathbf{g}(\mathbf{q}_j) = \boldsymbol{\tau} + J(\mathbf{q}_j)^\top \mathbf{f}_c, \tag{32}$$

where $\mathbf{q}_j$ gathers the joint coordinates, $\boldsymbol{\tau}$ denotes the actuation torques and $\mathbf{f}_c$ denotes the foot contact force. The abduction, hip pitch and knee joints are driven by random driving torques, each formed by three sinusoids with randomized amplitudes, frequencies and phases plus an Ornstein Uhlenbeck perturbation. The resulting joint reaction and contact histories serve as boundary loads for a componentwise linear elastodynamic model

$$\rho\,\ddot{\mathbf{u}} = \nabla\cdot\boldsymbol{\sigma} + \mathbf{b}(\mathbf{x}, t), \tag{33}$$

solved with FEniCSx using the Newmark average acceleration scheme at $\Delta t = 0.01$ s. The input series contains 17 channels, grouping joint positions, joint velocities, joint torques, foot reaction

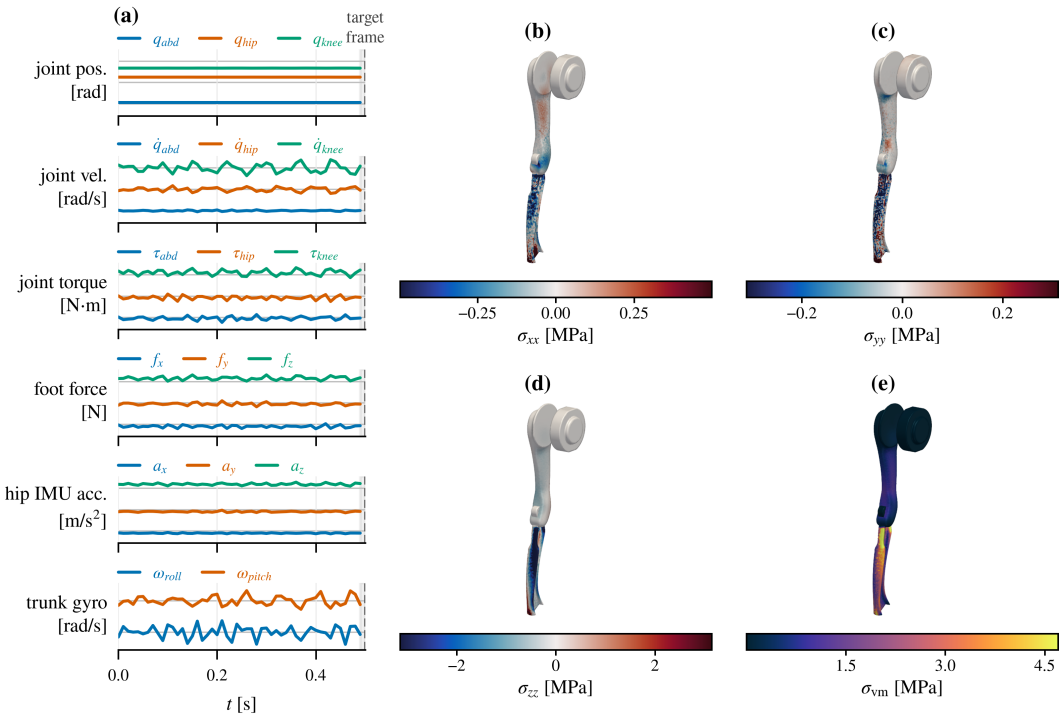

Figure 8: Go1Leg-Dyn. Panel a shows the 50 step input window of the 17 channel signal, grouped into joint positions in rad, joint velocities in $\mathrm{rad\,s^{-1}}$, joint torques in $\mathrm{N\,m}$, foot reaction forces in N, hip inertial accelerations in $\mathrm{m\,s^{-2}}$ and trunk gyroscope rates in $\mathrm{rad\,s^{-1}}$, with the dashed line marking the target frame. Panels b to e show the target frame stress field on the assembled leg mesh, including the normal components σ_{xx}, σ_{yy} and σ_{zz} in MPa with symmetric color scales, and the von Mises stress σ_{vm} in MPa, which concentrates near the knee joint with a peak of about 4.7 MPa for this sample. The displayed sample is the median field energy sample of the test split.

forces, hip mounted inertial accelerations and trunk gyroscope rates, which covers the actuation, kinematic and contact observables available on a real robot while remaining distinct from the internal stress field. The target field is the six component stress tensor over the assembled leg one step after the window. The stress concentrates near the knee joint and reaches 39 MPa over the dataset, reproducing the load transmission path of a real robot leg with a level of geometric and physical complexity beyond canonical benchmarks. Figure 8 illustrates one representative test sample.

C METRICS

Let $\bar{\boldsymbol{Y}}$ and $\hat{\bar{\boldsymbol{Y}}}$ denote the ground-truth and predicted fields after standardization with training-set statistics, and let $\boldsymbol{Y}, \hat{\boldsymbol{Y}}$ denote the corresponding fields in physical units. We report three complementary metrics. The normalized mean squared error

$$\mathrm{NMSE} = \frac{\|\hat{\bar{\boldsymbol{Y}}} - \bar{\boldsymbol{Y}}\|_F^2}{\|\bar{\boldsymbol{Y}}\|_F^2} \tag{34}$$

measures the global fit of the field energy in the normalized space and serves as the primary metric, comparable across datasets and channels of different scales. The relative L_2 error

$$\mathrm{RelL2} = \frac{\|\hat{\boldsymbol{Y}} - \boldsymbol{Y}\|_F}{\|\boldsymbol{Y}\|_F} \tag{35}$$

reports the same global fidelity in physical units, confirming that improvements in the normalized space are not artifacts of the standardization. To probe the worst-case behavior that governs engineering safety, each field is reduced to an equivalent scalar intensity s_i per node, the von Mises

equivalent stress for tensor fields and the vector magnitude for flux fields, and the node-wise relative errors $e_i = |\hat{s}_i - s_i|/(|s_i| + \varepsilon)$ with $\varepsilon = 0.01\,\langle s \rangle$ guarding against near-zero denominators give

$$\text{VM-MaxRE} = \max_i \; e_i, \tag{36}$$

the largest local relative error over the whole field. The three metrics are jointly necessary: NMSE and RelL2 reflect average performance and can mask localized failure, while VM-MaxRE isolates the tail error at critical regions such as stress concentrations, which is precisely where reliable prediction matters most for downstream structural assessment.

D Implementation Details

D.1 Unified Evaluation Protocol

All methods are evaluated under an identical protocol. Each dataset is split into $375/62/63$ training, validation, and test samples with a fixed split seed shared by every method, and normalization statistics are computed from the training subset only. Mesh cells are converted to bidirectional node-adjacency edges with full connectivity within each cell and duplicate edges removed; graph-based methods consume the edge list, while transformer and point-query methods consume the node point set. The training loss is the objective of Eq. equation 13 with $\lambda = 0.5$ in the standardized field space. Optimization uses Adam with learning rate 5×10^{-4}, weight decay 10^{-4}, cosine annealing, batch size 8, at most 200 epochs, early stopping with patience 30 on validation NMSE, and gradient clipping at 1.0. All reported numbers are averaged over five independent training seeds. Parameter counts and single-sample inference latency, measured as the median of one hundred forward passes, are recorded for every method.

D.2 Hyperparameter Tuning of Baselines

Every baseline is tuned before the final comparison, and the reported results correspond to its best validated configuration, so that each method is compared at its strongest rather than at an arbitrary default. The search is performed on the validation split with the unified training protocol above, over the following ranges:

- hidden width: $\{64, 128, 256\}$;
- depth: message-passing steps or attention layers in $\{4, 8, 10, 15\}$ for graph methods and $\{4, 6, 8\}$ for transformer methods;
- learning rate: $\{10^{-4}, 5 \times 10^{-4}, 10^{-3}\}$;
- method-specific structural hyperparameters: the POD truncation rank of SHRED in $\{16, 32, 64, 128\}$, the slice count of Transolver in $\{32, 64\}$, and the mixture-of-experts capacity of GNOT in $\{4, 8\}$ experts;
- the sensor embedding dimension is fixed at 128 for all methods to keep the injected conditioning information identical.

D.3 Baseline Implementations and Adaptations

This subsection describes, for each baseline, the implementation basis and the adaptation required by our task contract, under which no method may access any current or past field observation.

DeepONet (Lu et al., 2021). We use a standard branch–trunk implementation: the branch MLP encodes the sensor embedding $\boldsymbol{z}_s$ and the trunk MLP encodes node coordinates, and their inner product gives the field value at each node. The model queries nodes independently and receives no mesh topology.

S-DeepONet (He et al., 2024c). We follow the published design: a gated recurrent branch encodes the observation history and an MLP trunk encodes coordinates, merged by inner product. The recurrent encoder matches the hidden width of our temporal encoder.

SHRED (Williams et al., 2024). We retain the native architecture: a recurrent encoder maps the sensor history to coefficients of a POD basis, which are decoded by the frozen basis to the full field. The POD basis is computed by singular value decomposition of the training fields with the truncation rank tuned as described above. SHRED receives no coordinates or topology, consistent with its original formulation.

MeshGraphNets (Pfaff et al., 2021). We build on a widely used community PyTorch reproduction. Since the original model rolls out the previous field autoregressively, which our contract forbids, we remove the field input channels and instead broadcast the sensor embedding to every node, denoted as the adapted variant. The encode-process-decode structure, edge features, and message-passing depth follow the original work.

BSMS-GNN (Cao et al., 2023). We use the bi-stride multi-scale graph network with pooling depth three as in the original work. The same adaptation as MeshGraphNets is applied: the field history input is removed and replaced by the broadcast sensor embedding.

LSTM-GNN (Guevara Garban et al., 2026). We follow the published coupling of a recurrent observation encoder with a graph neural network: the native LSTM encodes the observation history into a hidden state that is broadcast to node features, followed by message passing on the mesh. The physics-informed loss terms of the original work are disabled because they are specific to its constitutive setting, and the model is trained with the unified objective; this is the adapted variant.

GNOT (Hao et al., 2023). We use the official architecture with its global conditioning slots set to the sensor embedding, so that the observed state information enters through the designed gating interface. Geometric input is given as point coordinates, and the mixture-of-experts cross-attention is retained.

Transolver (Wu et al., 2024). We use the official implementation with Physics-Attention over learnable slices. Node inputs are coordinates concatenated with the sensor embedding, and the slice count is tuned as described above.

Erwin (Zhdanov et al., 2025). We use the official ball-tree hierarchical attention implementation over ragged point sets, with the sensor embedding concatenated to point features. The tree depth and ball sizes follow the recommended configuration for meshes of our scale.

D.4 SONAR Implementation

The temporal encoder is a single-layer gated recurrent unit with hidden width $d_h = 128$, and the sensor embedding is its final hidden state. The conditional spectral basis generator uses node features given by coordinates concatenated with the broadcast embedding, edge features given by unit direction and length, $K = 15$ message-passing steps with unshared per-step parameters, residual connections, and layer normalization, followed by a two-layer per-node per-channel decoder producing $R = 64$ basis fields. The Gram-whitening layer is implemented through a Cholesky factorization of the Gram matrix with a jitter initialized at 10^{-4} and escalated by a factor of ten until the factorization succeeds. The coefficient head is a shallow MLP taking the sensor embedding and a globally pooled geometry embedding. The residual branch estimates curvature by the discrete Laplacian energy of the detached global field, selects the top 10% of nodes expanded by two hops as the active set, and applies four message-passing steps on the induced subgraph with node features given by coordinates, the sensor embedding, and the global field. On the Go1Leg-Dyn dataset the input dimension of the temporal encoder and of all baselines is adjusted to seventeen channels, with no other structural change. The full model has 2.4M parameters, in the same range as MeshGraphNets. When the total number of edges in a batch exceeds a memory threshold, message passing falls back to per-sample serial evaluation, which changes only the computation schedule and not the model semantics or gradients.

Table 5: Full comparison on the three benchmarks with per-seed variability. Entries are mean ± one standard deviation over the same five training seeds as Table 1; **bold** and underline mark the best and second-best *mean* per column. Improvement is SONAR's relative reduction against the strongest baseline, and p-values report two-sided Welch's t-tests across seeds (n=5 per method); both rows are repeated from Table 1 for convenience. Standard deviations below 10^{-3} are printed with additional digits. Metric definitions are given in Appendix C.

	ElasticBeam-Syn			HeatPlate-Syn			Go1Leg-Dyn		
Method	NMSE↓	RelL2↓	VM-MaxRE↓	NMSE↓	RelL2↓	VM-MaxRE↓	NMSE↓	RelL2↓	VM-MaxRE↓
DeepONet	0.3889 ± 0.0053	0.3048 ± 0.0034	4.549 ± 0.110	0.1373 ± 0.0108	0.2989 ± 0.0237	13.464 ± 0.311	0.8899 ± 0.0062	0.7963 ± 0.0054	346 ± 9.4
MeshGraphNets	$\underline{0.0263} \pm 0.00095$	$\underline{0.1606} \pm 0.0061$	$\underline{2.851} \pm 0.082$	$\underline{0.0007} \pm 0.000042$	$\underline{0.0318} \pm 0.0004$	0.784 ± 0.033	0.2390 ± 0.0051	0.4017 ± 0.0072	26.4 ± 2.2
BSMS-GNN	0.0469 ± 0.0018	0.1853 ± 0.0027	3.745 ± 0.098	0.0046 ± 0.000677	0.0562 ± 0.0019	1.577 ± 0.056	0.3094 ± 0.0064	0.4864 ± 0.0087	32.4 ± 2.9
GNOT	0.1113 ± 0.0086	0.2524 ± 0.0019	4.166 ± 0.169	0.0105 ± 0.0015	0.0762 ± 0.0081	3.001 ± 1.883	0.5823 ± 0.0118	0.5444 ± 0.0096	102 ± 13.7
SHRED	0.0397 ± 0.0010	0.2187 ± 0.0058	3.043 ± 0.192	0.0023 ± 0.000155	0.0354 ± 0.0014	0.763 ± 0.010	0.2522 ± 0.0016	0.4739 ± 0.0019	$\underline{13.3} \pm 0.64$
S-DeepONet	0.3879 ± 0.0038	0.3010 ± 0.0060	4.370 ± 0.131	0.1275 ± 0.0071	0.2774 ± 0.0166	12.965 ± 0.209	0.8393 ± 0.0050	0.7094 ± 0.0063	372 ± 8.1
Transolver	0.3625 ± 0.0041	0.2816 ± 0.0036	5.545 ± 0.370	0.0252 ± 0.0020	0.1273 ± 0.0045	7.303 ± 1.185	0.4438 ± 0.0089	0.4876 ± 0.0081	87.5 ± 9.3
Erwin	0.0506 ± 0.0097	0.2036 ± 0.0257	3.863 ± 0.357	0.0056 ± 0.000473	0.0570 ± 0.0023	1.792 ± 0.142	0.3398 ± 0.0152	0.4537 ± 0.0179	169 ± 24.6
LSTM-GNN	0.0604 ± 0.0026	0.2514 ± 0.0110	3.939 ± 0.374	$\underline{0.0007} \pm 0.000028$	0.0333 ± 0.0010	$\underline{0.742} \pm 0.0060$	$\underline{0.2122} \pm 0.0045$	$\underline{0.3930} \pm 0.0148$	43.9 ± 3.6
SONAR (Ours)	$\mathbf{0.0249} \pm 0.00089$	$\mathbf{0.1600} \pm 0.0052$	$\mathbf{2.730} \pm 0.070$	$\mathbf{0.0005} \pm 0.000084$	$\mathbf{0.0281} \pm 0.0004$	$\mathbf{0.734} \pm 0.0051$	$\mathbf{0.1897} \pm 0.0040$	$\mathbf{0.3580} \pm 0.0126$	$\mathbf{11.7} \pm 0.55$
Improvement	5.3%	0.4%	4.2%	28.6%	11.6%	1.1%	10.6%	8.9%	12.0%
p-value	0.043	0.872	0.037	0.004	<0.001	0.054	<0.001	0.004	0.003

E FULL RESULTS

Table 5 expands the main comparison with per-seed standard deviations. All entries are mean $\pm$ one sample standard deviation over the same five training seeds, and the data splits, training budgets, and metric definitions in Appendix C are identical to the main text. The test split is fixed across seeds, so the reported dispersion reflects training stochasticity rather than test-set sampling noise. Following the main-text protocol, GAOT, GINO, and UPT are omitted because they diverge or collapse to mean prediction on these benchmarks. The p-values are two-sided Welch's t-tests between SONAR and the strongest baseline of each column. We use Welch's test rather than a rank-based alternative because with five seeds the smallest attainable two-sided Wilcoxon p-value is 0.0625, which cannot resolve the effects reported here.

Three patterns emerge from the dispersion structure. First, absolute spreads contract sharply from Go1Leg-Dyn to ElasticBeam-Syn to HeatPlate-Syn. The smooth and low-rank HeatPlate-Syn field makes training near-deterministic for the leading methods, whereas the contact-rich transients of Go1Leg-Dyn induce the largest seed sensitivity, most visibly in Erwin's NMSE spread of ± 0.0152 and GNOT's VM-MaxRE spread of ± 13.7. Second, VM-MaxRE is consistently the noisiest metric for all methods, since a maximum over thousands of nodes is dominated by a handful of stress-concentration nodes whose predictions vary with the training trajectory. Third, and most relevant to our claims, SONAR attains the smallest VM-MaxRE standard deviation among all ten methods on all three benchmarks, at 0.070, 0.0051, and 0.55 respectively, so its worst-case gains are systematic rather than the product of a favourable seed.

On ElasticBeam-Syn, SONAR pairs the best mean with the smallest NMSE spread of ± 0.00089, while Erwin, the strongest hierarchical-attention baseline on this benchmark, fluctuates at ± 0.0097 with a coefficient of variation of 19.2%, and its one-standard-deviation band overlaps the mean of clearly weaker methods. Across all nine columns, SONAR's spread is smaller than or comparable to that of the strongest baseline in eight cases, and the sole exception on HeatPlate-Syn NMSE concerns absolute spreads below 10^{-4}, where relative comparisons cease to be meaningful. We attribute this joint accuracy and stability profile to the conditioned and Gram-whitened spectral basis: constraining the basis geometry across seeds confines run-to-run variation to the low-dimensional coefficient head, which is consistent with the loss-landscape analysis in the main text. Read through the lens of dispersion, the conclusions of Table 1 are therefore unchanged but better qualified: seven of the nine improvements are significant at $p < 0.05$, the single statistical tie on ElasticBeam-Syn RelL2 and the single marginal case on HeatPlate-Syn VM-MaxRE both sit well inside the seed-noise band, and the significant improvements are accompanied by equal or smaller seed dispersion in essentially all columns.

F USE OF LARGE LANGUAGE MODELS

During the preparation of this paper, we made limited and controlled use of large language models (LLMs) in the following two aspects.

Writing aid and polishing. LLMs were used solely as auxiliary language-editing tools to improve the clarity, grammar, fluency, and readability of text originally drafted by the authors. No scientific statements, claims, or results were generated or altered by LLMs beyond stylistic refinement.

Retrieval and discovery. LLMs were used to aid literature retrieval and discovery, such as identifying potentially relevant related work and assisting with keyword-based searches. All identified references were subsequently located, read, and verified in their original sources by the authors, and only references directly consulted by the authors are cited in this paper.

All core scientific contributions of this work — including the research problem formulation, the theoretical analysis and derivations, the algorithmic and architectural design, the experimental protocols, and the interpretation of results — were independently conceived, developed, implemented, and verified by the authors. LLMs did not contribute to the intellectual or scientific content of this work.