RELATIONFIRST: WHAT COMPOSITION REQUIRES FROM LEARNED REPRESENTATIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

Accurate local predictions do not establish whether a learned representation will remain reliable within a larger computation. The open question is what information the consuming computation needs and how errors in that information affect its result. RelationFirst makes both requirements precise. Composition sufficiency identifies the distinctions a component must preserve across contexts, while propagation geometry determines how local approximation errors affect the assembled result. Stationary elliptic PDEs make these requirements testable. Classical elimination exposes an exact interface for the discrete problem, and the assembled equilibrium reveals how response errors propagate. We train models on isolated components, freeze them, and reuse them in held-out assemblies. With the output decoder and composer fixed, changing only the supplied interaction shifts the global equilibrium. A single propagation law predicts these effects across distinct nonlinear systems, including cases where a learned response helps and where it harms. Perhaps counter-intuitively, in a controlled intervention, even a local error with only half the magnitude yields a larger global error across all 72 tested assemblies, driven purely by the shift in error direction. These results show why reusable representations must be evaluated through the behavior they preserve under composition and the consequences of their errors in use. These criteria can guide the design and evaluation of learned components in numerical solvers and coupled simulators.

1 INTRODUCTION

We study a representation problem under composition. A learned module is usually evaluated on the output it was trained to produce. Once reused within a larger computation, it must also supply the interaction that coordinates it with other modules. In neural PDEs, a local predictor returns a field, while the surrounding assembly needs a boundary response. *Does output fidelity alone certify reliable compositional reuse?*

Elliptic PDEs let us identify the required interface. Classical elimination gives the boundary response needed to assemble an elliptic system. This response is known independently of learning. We can therefore study what the representation contains, how accurately it is evaluated, and how the field is reconstructed. Models trained on isolated component queries are frozen before reuse in unseen assemblies. We then test which distinctions surrounding components need and how response errors change the assembled solution.

We isolate the resulting *output–reuse mismatch* while holding the learned output decoder and global composer fixed. Only the local response evaluator and its consistent tangent change. This is the first controlled identification of the output–reuse mismatch in nonlinear elliptic composition.

A fixed decoder can produce very different solutions. With each decoder frozen, HF responses reduce mean energy-norm error from 3.777% to 2.081% in cubic and from 0.208% to 0.097% in sinh (Figure 1). Every case improves for each of the three decoders in both families. A dedicated learned relation closely matches the HF-response result in cubic. The current learned sinh relation increases field error to 1.406%. Both systems need the same kind of interface, and the learned evaluators supply it with very different accuracy.

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.

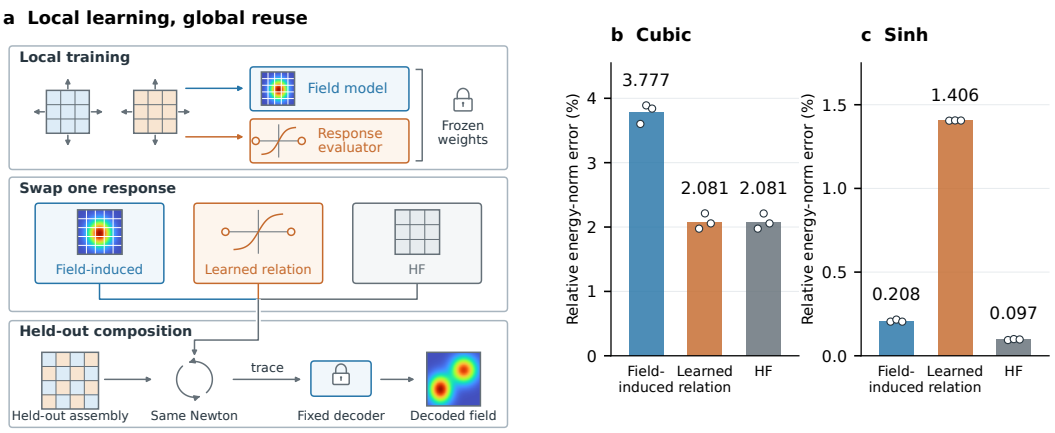

Figure 1: **The same output model changes its behavior under composition.** **a**, isolated component queries train the field model $\mathcal{D}$ and the dedicated response evaluator. Frozen weights enter held-out compositions. One of three response evaluators supplies the same Newton solver, and the same $\mathcal{D}$ decodes its equilibrium trace. Icons illustrate the protocol. **b,c**, mean relative energy-norm errors for cubic and sinh, with three decoder seeds per family. Dots are seed means. Cases first average nodes17/33, then cases and seeds receive equal weight. Cubic uses 32 cases and sinh 16, with one shared relation per family. References use nonlinear HF reconstruction on the same mesh and 16-dimensional port. Protocols are in Appendices A and A.6.

The theory specifies what to preserve and how errors propagate. Composition sufficiency identifies what a surrounding system can detect about a component. Approximation errors then meet at shared boundaries, where they may cancel or reinforce one another. The assembled equilibrium turns the remaining error into a global displacement. We derive this error geometry and finite-error bounds.

One propagation law describes both outcomes. It predicts equilibrium displacement in both families, with median relative prediction errors below 1% for all six field models and both dedicated relations (Table 1). A separate directional intervention isolates this mechanism on 72 assemblies. An error with half the local magnitude produces greater global error on every one.

We make three contributions. **First**, we identify the output–reuse mismatch through a frozen-decoder intervention in two nonlinear elliptic families, each with three decoder seeds. **Second**, we formalize composition sufficiency and derive a propagation law with finite-error bounds, linking the information an interface preserves to the effect of its errors. **Third**, we test this law across cubic and sinh and isolate error direction as a cause of reversed local and global accuracy rankings.

2 The Output–Reuse Mismatch

2.1 Separating Prediction from Interaction

Consider a component K with parameters p_K, boundary trace λ_K, and interior field u_K. A frozen model $\mathcal{D}_K(\lambda_K, p_K)$ predicts the field. The surrounding assembly consumes an outward boundary response t_K. Shared traces and balanced responses determine the equilibrium at which the field model is evaluated.

We study two nonlinear reactions in

$$-\nabla\cdot(a\nabla u) + cu + \gamma\phi(u) = f, \qquad \phi(u) \in \{u^3, \sinh u - u\}, \tag{1}$$

with $a, c > 0$ and $\gamma \geq 0$. The domain contains 4×4 square components. Local Q1 finite elements use 17 or 33 nodes per axis and a fixed 16-dimensional trace space. HF means high fidelity and NN means neural network. HF solves the local nonlinear discrete equations to numerical tolerance.

Three field decoders per family impose traces exactly through a linear lifting and a bubble factor. They train on local fields with an interior residual penalty. Dedicated response models also train on

isolated probes. Each family has its own data and weights, frozen before assembly (Appendices A and A.6).

2.2 One output model, three response evaluators

The *field-induced response* evaluates the discrete weak boundary residual on $\mathcal{D}_K(\lambda_K, p_K)$. Its tangent differentiates the complete field-to-response map. The *learned relation* supplies an independently trained local response. The *HF response* equilibrates the local interior before returning its response and tangent. Each enters the same assembly and Newton code. After convergence, every arm reconstructs with the unchanged $\mathcal{D}_K$ at its own trace.

Held fixed Field-model weights and reconstruction, physical inputs, mesh, port, initialization, and global Newton solver.

Exchanged Local response and its consistent tangent.

A test case fixes the layout, coefficients, forcing and boundary data. Cubic uses 32 held-out cases, eight each for checkerboard, inclusion, repeated-block and stripe layouts, shared by the decoder replications. Sinh uses 16 held-out cases. We average meshes within each case and weight cases equally within each family.

Metrics. Relative L^2 error measures field values. The energy norm also measures gradients (Figure 1). Relative interface-trace error measures equilibrium in the R-norm. References use same-mesh, same-port HF (Appendix A.7).

2.3 The response affects what the decoder predicts

A better interaction improves the same decoder in both families. HF responses improve every case for each decoder in both families (Figure 1), reducing mean energy-norm error by 44.9% in cubic and 53.4% in sinh. The remaining field error is reconstruction error at the reference trace. Relative L^2 error falls from 3.491% to 0.536% in cubic and from 0.160% to 0.0206% in sinh, with every case improving for each decoder (Table 10).

In cubic, the learned relation gives 2.081% mean energy-norm error. All nine pairings of three decoders and three recorded relation checkpoints improve every case, with geometric mean error ratio 1.824 and 95% bootstrap interval [1.683, 1.969] (Appendix A.9).

In sinh, field-induced and learned-relation interface errors are 0.190% and 1.450%. Response substitution increases the fixed decoder's field error from 0.208% to 1.406%. All three response choices converge without HF fallback. The models have different training protocols and capacities. Their responses enter the same composer.

Adding response and Jacobian–vector-product (JVP) supervision to the cubic field model does not materially reduce its interface error. The geometric mean error ratio is 1.024, with a 95% interval of [0.940, 1.115] (Appendix A.8).

Composition chooses the inputs at which prediction is used. Local validation evaluates $\mathcal{D}$ at prescribed traces. Composition determines its input $\widehat{\Lambda}$ at equilibrium. Relative to the HF trace $\Lambda^\star$ and field $u^\star$,

$$\mathcal{D}(\widehat{\Lambda}) - u^\star = \big[\mathcal{D}(\widehat{\Lambda}) - \mathcal{D}(\Lambda^\star)\big] + \big[\mathcal{D}(\Lambda^\star) - u^\star\big].$$

Fixing $\mathcal{D}$ preserves the second term. Changing the response evaluator moves its equilibrium input and changes the first term.

3 COMPOSITION-SUFFICIENT REPRESENTATIONS

To explain the output–reuse mismatch, we first specify what the assembly needs to know about each component. Composition sufficiency formalizes this information requirement. Exact elliptic elimination gives a concrete sufficient interface. Section 4 then studies the effect of approximating it.

3.1 COMPOSITION SUFFICIENCY

Let X be a component, C an admissible surrounding context, and $O(C[X])$ a declared observable of the composed system. The context includes the coupling rules, admissible inputs, and solution branch. Two components are contextually equivalent when

$$X_1 \sim_{\mathcal{C},O} X_2 \quad \Longleftrightarrow \quad O(C[X_1]) = O(C[X_2]) \quad \text{for every } C \in \mathcal{C}. \tag{2}$$

Definition 1 (Composition sufficiency)**.** A representation r is sufficient for $(\mathcal{C}, O)$ if $r(X_1) = r(X_2)$ implies $X_1 \sim_{\mathcal{C},O} X_2$. An executable realization also supplies a composer that obtains the declared observable from the represented components.

Two components in the same contextual class behave identically in every allowed composition. The quotient $\pi(X) = [X]_{\mathcal{C},O}$ records this class. A sufficient representation may preserve more, but it must determine the quotient, so $\pi = h \circ r$ for some h. An *exact contextual representation* also maps equivalent components to the same value. Appendix B.2 proves this factorization and specifies its informational meaning. The elliptic result below establishes sufficiency for a fixed trace space. Port capacity is examined separately.

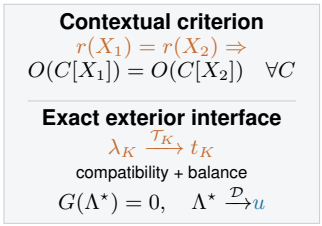

Figure 2: **Composition sufficiency.**

3.2 THE BOUNDARY RESPONSE IS SUFFICIENT FOR ELLIPTIC ASSEMBLY

Fix a component discretization and split its variables into exterior traces λ_K and interior variables w_K. On a regular branch, the interior equations $F_{K,I}(\lambda_K, w_K) = 0$ define a unique smooth solution $w_K^\star(\lambda_K)$. Elimination gives the external response

$$\mathcal{T}_K(\lambda_K) = F_{K,E}(\lambda_K, w_K^\star(\lambda_K)), \qquad \lambda_K = E_K \Lambda + \lambda_{K,D}. \tag{3}$$

Here E_K restricts the free global trace Λ to component K, and $\lambda_{K,D}$ contains prescribed exterior data. The response is a covector paired with trace variations. Adjacent components share traces and their outward responses sum to zero. Thus global equilibrium is

$$G(\Lambda) = \sum_K E_K^\top \mathcal{T}_K(E_K \Lambda + \lambda_{K,D}) = 0. \tag{4}$$

Static condensation (Eftang & Patera, 2014) and the implicit-function theorem (Bonnans & Shapiro, 2000) yield the following sufficiency result.

Proposition 1 (Sufficiency of the boundary response)**.** *Fix the discrete trace spaces and a branch on which every local interior solution is unique and has an invertible interior Jacobian. The roots of (4), together with the eliminated local interiors, correspond exactly to equilibria of the full discrete assembly on that branch. Equal local response maps on the admissible trace domain preserve all exterior equilibrium observables under the same coupling rules.*

The proof reinstates the eliminated interiors (Appendix B). This exact interface lets us exchange evaluators while preserving what a component must communicate. Port capacity sets what can be represented. Evaluator accuracy sets how faithfully it is supplied.

3.3 A RELATION, ITS EVALUATOR, AND ITS OUTPUT DECODER

The relation $\mathcal{T}_K$ represents interaction. Its evaluator supplies values and tangents. A decoder or local solve recovers interior fields from the assembled trace (Figure 2). The intervention changes the evaluator while keeping $\mathcal{D}_K$ fixed.

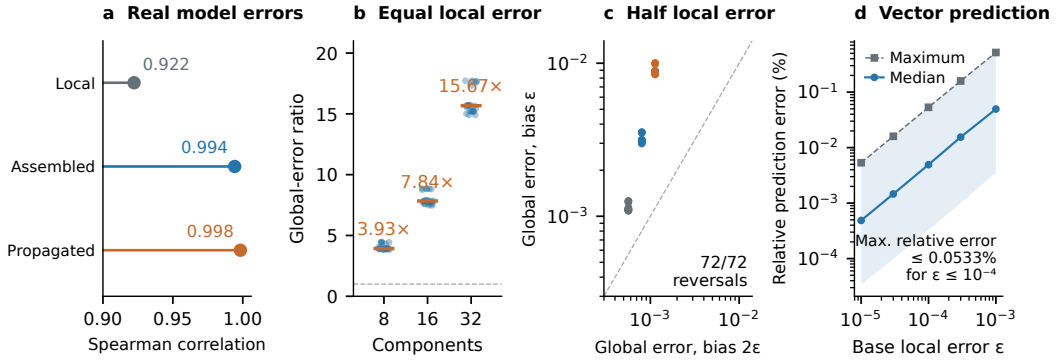

Figure 3: **The propagation law predicts equilibrium shifts and accuracy reversals.** **a**, actual field-induced errors from three decoders. Correlations with nonlinear equilibrium displacement use 96 seed–case pairs, with two meshes averaged in each pair and 32 shared test cases. **b–d**, a separate 1D intervention uses 72 assemblies and constant response biases with unchanged tangents. **b**, at $\epsilon = 10^{-4}$, equal-norm sensitive and median singular directions give median nonlinear displacement ratios of 3.93, 7.84, and 15.67 for $N = 8, 16, 32$. **c**, the sensitive bias has half the local magnitude of the doubled-median bias, yet gives larger global error on all 72 cases. The line marks equal global error. **d**, relative prediction error $\|\delta y + Mq\|_2 / \|\delta y\|_2$, in percent. Each case takes its worst of three directions. Solid and dashed curves show the median and maximum across cases, with the full range shaded. The annotation summarizes the maximum measured error at base $\epsilon \leq 10^{-4}$. All 1,080 perturbed solves converge. Appendix C gives every magnitude and case.

4 Composition Changes the Geometry of Error

Once we know what a component must communicate, which errors in that communication matter? A local norm measures their size. The assembly also responds to where they occur and which way they point. We derive this dependence, measure it on trained evaluators, and isolate it experimentally.

4.1 The Propagation Map

Let $\Lambda^\star$ solve the exact assembly and let R be a positive-definite metric on its free traces. Write $y = R^{1/2}\Lambda$ and $\bar{G}(y) = R^{-1/2}G(R^{-1/2}y)$. The normalized Jacobian at equilibrium is $\bar{J} = R^{-1/2}JR^{-1/2}$, where $J = DG(\Lambda^\star)$. Let $\delta t = Lq$ stack local response biases in coordinates for which $\|q\|_2$ is the selected local error norm. The assembly matrix A sums these outward covectors on free interfaces. Define

$$B = R^{-1/2}AL, \qquad M = \bar{J}^{-1}B. \tag{5}$$

A standard implicit-function perturbation argument (Bonnans & Shapiro, 2000; Deuflhard, 2004) gives the local response of the equilibrium.

Proposition 2 (Error geometry at a regular equilibrium). *If $\bar{G}$ is twice continuously differentiable near $y^\star$ and $\bar{J}$ is invertible, the nearby root of $\bar{G}(y) + Bq = 0$ satisfies*

$$\delta y = -Mq + O(\|q\|_2^2), \qquad \|\delta y\|_2 = \left\|\widehat{\Lambda} - \Lambda^\star\right\|_R. \tag{6}$$

The assembled system weights error directions differently. The leading squared global error is $q^\top M^\top M q$. Local error measured by $q^\top q$ treats all directions equally. Composition weights them through $M^\top M$. The two quadratic forms agree up to a scalar only in the special case that this matrix is proportional to the identity on the relevant subspace. Errors in $\ker B = \ker M$ cancel during assembly, so $\|Mq\|_2$ is generally a seminorm on local errors. On the surviving subspace, the singular values of M determine their downstream weight. This directional view complements residual-based and goal-oriented error estimation (Becker & Rannacher, 2001). Here it also provides a direct intervention on the mechanism. Appendix C proves a finite-neighborhood bound and extends the remainder estimate to trace-dependent learned errors.

4.2 THE PROPAGATION LAW PREDICTS EQUILIBRIUM SHIFTS IN BOTH PDE FAMILIES

At the HF equilibrium of each 2D assembly, we evaluate the actual local defects $\delta t_K = \widehat{\mathcal{T}}_K(\lambda_K^\star) - \mathcal{T}_K^{\mathrm{HF}}(\lambda_K^\star)$. They give an assembled defect $d = \sum_K E_K^\top \delta t_K$ and the first-order perturbation prediction $-J_{\mathrm{HF}}^{-1} d$ of the nonlinear trace displacement. Assembly first combines the local defects. The inverse HF Jacobian then turns their sum into a trace displacement, exactly as the map M predicts.

In cubic, across the three field decoders, Spearman correlation with measured displacement is 0.922 for ordinary local dual error, 0.994 for $\|d\|_{R^{-1}}$, and 0.998 for $\left\|J_{\mathrm{HF}}^{-1} d\right\|_R$ (Figure 3a). Each seed–case pair first averages its two meshes. Most of the gain comes from assembling the local defects. Equilibrium sensitivity then improves their ordering further. The median relative prediction error is 8.13% for the field-induced interfaces and 0.0238% for the fixed learned relation. Appendix C.3 gives the distributions, per-seed results, and the larger finite-error deviations.

State-dependent defects also change the tangent. For an actual evaluator, write $\widehat{\bar{G}}(y) = \bar{G}(y) + \bar{d}(y)$. Its consistent Jacobian at the HF root is $\widehat{\bar{J}} = \bar{J} + D\bar{d}(y^\star)$. When this Jacobian is invertible and the nearby branch is regular, expansion of the candidate response system gives

$$\delta y = -\widehat{\bar{J}}^{-1} \bar{d}(y^\star) + O(\|\delta y\|_2^2). \tag{7}$$

Appendix C.2 states the Lipschitz conditions and proves the quadratic remainder bound. Equation (7) accurately predicts displacement for all eight evaluators across cubic and sinh (Table 1). The dedicated relations give median relative prediction errors of 0.000577% in cubic and 0.115% in sinh. The learned response improves field accuracy in cubic and worsens it in sinh (Figure 1). The law predicts the equilibrium shift in both cases. Constant injected biases have $D\bar{d} = 0$, recovering the geometry tested next.

Table 1: **One propagation mechanism across physics and evaluator quality.** Relative prediction errors (%) use observed displacement as denominator and the evaluator's consistent tangent at the HF equilibrium. Ranges span three seed medians after mesh pairing over 32 cubic or 16 sinh cases. Each family shares one dedicated relation. Appendix C gives distributions and HF-tangent comparisons.

PDE family	Response evaluator	Models	Median prediction error (%)
Cubic	Field-induced	3	0.833–0.983
Cubic	Dedicated relation	1	0.000577
Sinh	Field-induced	3	0.0051–0.0094
Sinh	Dedicated relation	1	0.115

4.3 CHANGING ERROR DIRECTION WHILE PRESERVING ITS MAGNITUDE

We use 72 existing one-dimensional cubic assemblies, comprising $N = 8, 16, 32$ components, four coefficient patterns, and six boundary replicates. This cohort is distinct from the 2D confirmation in Section 2. For each HF equilibrium, R is the assembled linear-diffusion energy matrix and the local norm is fixed as

$$\eta_0^2 = \frac{1}{N} \sum_K \frac{h_K}{a_K} \|\delta t_K\|_2^2, \qquad q_K = \sqrt{\frac{h_K}{N a_K}}\, \delta t_K. \tag{8}$$

Thus equal $\|q\|_2$ means equal stacked local RMS error. We compute the sensitive right singular vector $v_{\max}$ and the median nonzero right singular vector v_{mid} of M. At each of five fixed magnitudes, we inject $q = \epsilon v_{\max}$, $q = \epsilon v_{\mathrm{mid}}$, and $q = 2\epsilon v_{\mathrm{mid}}$ into the actual local HF responses. These trace-independent biases leave every local tangent unchanged. The same nonlinear solver is restarted from the corresponding HF equilibrium.

Directions are fixed from the reference geometry before any perturbed solve. We compare the observed displacement with $-Mq$ in both magnitude and direction.

4.4 SMALLER LOCAL ERROR CAN GIVE LARGER GLOBAL ERROR

The local ranking reverses on every tested assembly. At $\epsilon = 10^{-4}$, the sensitive direction gives larger global error on all 72 cases. At equal local error, the median global-error ratio rises from 3.93

at eight components to 15.67 at 32 (Figure 3b). These are within-assembly direction ratios, so each comparison holds the physical system and the local error budget fixed.

The stronger test reverses the local ranking. If $\sigma_{\max} > 2\sigma_{\mathrm{mid}}$, first-order propagation predicts that $\epsilon v_{\max}$ gives larger global error than $2\epsilon v_{\mathrm{mid}}$. All 72 measured spectra satisfy this condition, and all 72 nonlinear comparisons exhibit the predicted reversal (Figure 3c). The median global-error ratio is 3.92 despite a local-error ratio of one half.

The explanation is quantitatively accurate. For base magnitudes $\epsilon \leq 10^{-4}$, the worst relative prediction error $\|\delta y + Mq\|_2 / \|\delta y\|_2$ is 0.0533%, including all three bias conditions. Over the complete sweep through 10^{-3}, its maximum is 0.5181%. All 1,080 perturbed solves converge. The controlled intervention isolates the effect of direction at fixed local norm. For trained evaluators, assembly and equilibrium sensitivity explain how response errors affect the global solution.

5 Design Consequences of RelationFirst

The **RelationFirst principle** combines the two theoretical requirements. Preserve the distinctions visible to admissible compositions and measure errors through their downstream consequences. This separates three design choices.

5.1 The representation determines what can be communicated

The port fixes the traces that neighboring components can exchange. Refining the local mesh improves the interior discretization while leaving this communication space unchanged. In the 65-node cubic transfer, the frozen evaluator's maximum same-port trace error is 0.0981%. Yet only 12 of 16 reconstructed fields have energy-norm error below 1% against finer FEM. On a separate eight-case audit, enriching the port to 24 or 32 coefficients with HF or a third-order analytic response (Analytic-3) brings every field within 1% (Figure 4). Accurate evaluation and adequate interface capacity address different errors.

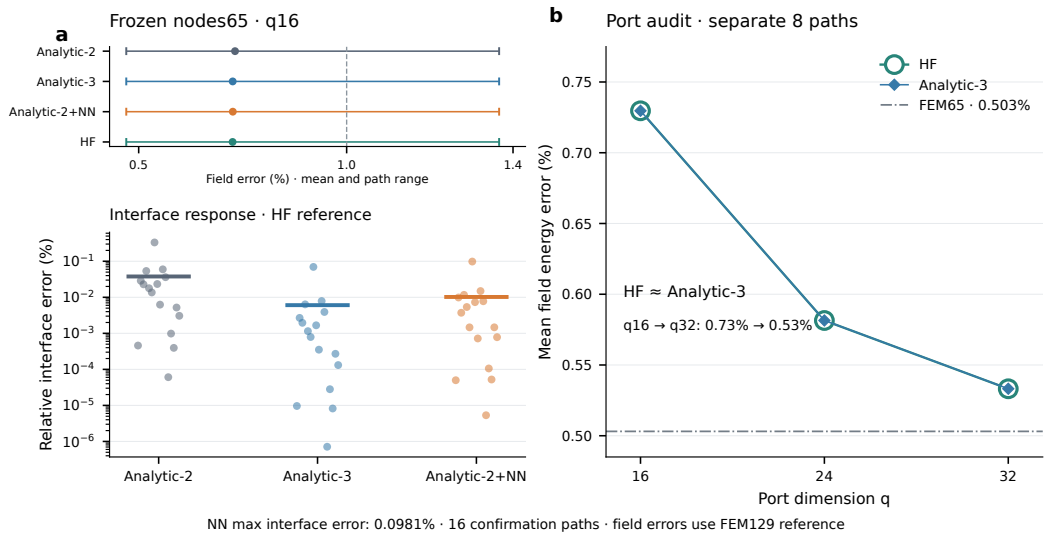

Figure 4: **Mesh refinement and port enrichment change different information.** Left, frozen cubic evaluation at 65 nodes per axis with a 16-dimensional port, over 16 cases. Right, port enrichment on eight cases. Relative trace errors use same-port HF. Energy-norm errors use finer FEM. Norms and references are in Appendix E.3.

5.2 THE EVALUATOR DETERMINES HOW ACCURATELY THE INTERFACE IS SUPPLIED

Once the boundary response is chosen as the interface, it can be evaluated with a field predictor, an analytic expansion, a dedicated network or an HF solve. Our main learned evaluator uses

$$\widehat{\mathcal{T}}_K = \mathcal{T}_K^{[2]} + \omega(s_K) s_K Q_\theta^{\mathrm{phys}}. \tag{9}$$

The analytic response supplies the known part. A physics-derived scale s_K normalizes the unresolved remainder, and ω is a calibrated activation. We call this evaluator *Analytic-2+NN*. Analytic-2 and Analytic-3 retain second- and third-order analytic response expansions. The primary instantiation uses a second-order split. The framework permits other orders. Appendix D gives the formulas and training details.

The cubic comparison shows how much the parameterization of this response matters. At 33 nodes, a direct response, raw remainder and normalized remainder give 32.1%, 15.0% and 0.00959% mean interface error. The value-only normalized control gives 0.00571%. Appendix E.1 specifies their training objectives. Their ordering motivates separate evaluation of response values, consistent tangents and composed outcomes.

Table 2: **Different evaluators approximate the same boundary response.** Relative interface-trace errors (%) against same-port HF, averaged over two meshes per case. Cubic has 32 cases, elasticity and sinh 16 each. The learned column reports mean ± sample SD of three seed means. Families use their own weights. Elasticity includes recorded HF fallback.

PDE	Analytic-2	Analytic-3	Analytic-2+NN	HF
Cubic	0.036	0.00566	0.0087 ± 0.0013	Reference
Elasticity	0.431	0.0018	0.175 ± 0.0085	Reference
Sinh	3.53	0.256	1.45 ± 0.055	Reference

Across three elliptic families, the learned remainder improves the analytic response it augments, while Analytic-3 gives the smallest interface errors (Table 2). In elasticity, 15 of 16 cases use no fallback. All 14 fallback calls per seed occur on one stripe case across two meshes (Appendix E.2).

5.3 THE DECODER DETERMINES THE REQUESTED OUTPUT

A response determines the exterior equilibrium. A decoder supplies the requested interior field at that equilibrium. Figure 1 makes the two errors separately observable by keeping the decoder fixed and supplying an HF trace. The port audit identifies a further restriction imposed by the trace space. Keeping these choices separate reveals which approximation controls the final output.

Eliminating a group's internal interfaces yields another exterior relation on a regular branch (Martinsson, 2015). Appendix F develops this closure and reports parent materialization and repeated-response costs, showing the computational consequences of retaining an executable external interface.

6 RELATED WORK

Behaviorally sufficient and compositional representations. Representation learning has long selected information by its downstream behavioral role. Bisimulation-based control representations preserve reward and transition behavior while discarding irrelevant observation detail (Zhang et al., 2021). Predictive-state representations describe dynamical state through action-conditional predictions of future observations (Littman et al., 2001). Relational inductive biases make entities and interactions explicit (Battaglia et al., 2018). Modular interfaces support generalization to unseen combinations of observation, action, and instruction spaces (Luketina et al., 2025). The success of compositional representations also depends on the learning rule and data (Lippl & Stachenfeld, 2025). RelationFirst studies what a predictor must supply when it becomes part of a larger computation. Elliptic elimination gives the required interface independently of learning, so we can change its evaluation while keeping the output decoder fixed. We then test how those changes alter the equilibrium.

Local neural models and numerical composition. Field and operator learning amortize requested solution maps (Raissi et al., 2019; Li et al., 2021; Lu et al., 2021; Kovachki et al., 2023). Learned DtN maps already approximate the local external response of nonlinear elliptic PDEs (Boutilier et al., 2025). SNI and NOEM combine local operator learning with domain decomposition and finite-element assembly (Huang et al., 2026; Ouyang et al., 2026). Principal-part decomposition learns a DtN remainder (Ling et al., 2026). Related constructions learn multiscale spaces, hybridizable discretizations, and local operators for system-level analysis (Rudikov et al., 2026; Haines & Shi, 2026; Fabiani et al., 2026).

The Convex Neural Energy Elements approach connects the learned object to assembly. Accurate fields can induce unsuitable Hessians, motivating convex energies with physical nullspaces for stable monolithic assembly (Jiang et al., 2026). RelationFirst asks which boundary information is sufficient for composition and how errors in that information propagate. Its response model permits nonsymmetric tangents and need not be derived from an energy potential.

Static condensation and port reduction provide the elliptic interface (Eftang & Patera, 2014; Smetana & Patera, 2016; Discacciati & Hesthaven, 2024). Implicit-function analysis supplies the root expansion (Bonnans & Shapiro, 2000; Deuflhard, 2004). Our contribution links what the interface preserves to how its errors propagate, then tests both with the decoder and composer fixed.

7 Discussion

From outputs to reusable representations. Local field accuracy does not fix the trace a decoder receives after assembly. Our frozen-decoder experiments show this gap in both PDE families. A learned response helps in one and harms in the other, while the same propagation mechanism predicts both shifts. Direction alone can make a smaller local error produce a larger global one.

Two requirements for representation learning. Predictive-state and behavioral representations select information by its role in future prediction and control (Littman et al., 2001; Zhang et al., 2021). Relational and modular learning organize reusable interactions (Battaglia et al., 2018; Luketina et al., 2025). RelationFirst adds an approximation requirement. Composition sufficiency determines what must be exposed. Propagation geometry determines which errors the consumer amplifies.

A physical laboratory for representation learning. Physical systems can reveal whether a learned representation encodes governing parameters (Qu et al., 2026). Here, elliptic physics also tells us what neighboring components must exchange. We can calculate that interface before learning, then change its evaluator, hold the output decoder fixed, and measure the resulting equilibrium shift. Models trained on isolated components face held-out compositions with frozen weights. This gives representation hypotheses a physically governed test with mechanistic control and quantitative prediction.

Evaluating representations in composition. Evaluation should specify the contexts of reuse, the observable behavior to preserve, and the consequences of approximation within those contexts. Local-validation and composed-error rankings reverse in all three tested training trajectories, with modest selection losses (Appendix A.10). This ranking reversal makes propagation geometry relevant to checkpoint choice. Unknown interfaces and history-dependent interactions require further identification of the information that reuse demands.

A reusable representation is an interface to future computations. We therefore need to evaluate both the information it makes available and the changes its errors cause when other components depend on it.

REPRODUCIBILITY STATEMENT

The supplement gives study protocols, training configurations, references, statistical units and mathematical assumptions. The source package contains recorded rows, plotting code and controlled-intervention matrices for independent reaggregation. Original confirmation and subsequent diagnostics are identified separately.

AI USE STATEMENT

AI tools assisted manuscript organization, language editing, LaTeX preparation, schematic illustration, figure production from recorded data, and consistency checking.

REFERENCES

Peter W. Battaglia, Jessica B. Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, Caglar Gulcehre, Francis Song, Andrew Ballard, Justin Gilmer, George Dahl, Ashish Vaswani, Kelsey Allen, Charles Nash, Victoria Langston, Chris Dyer, Nicolas Heess, Daan Wierstra, Pushmeet Kohli, Matt Botvinick, Oriol Vinyals, Yujia Li, and Razvan Pascanu. Relational inductive biases, deep learning, and graph networks, 2018. URL https://arxiv.org/abs/1806.01261.

Roland Becker and Rolf Rannacher. An optimal control approach to a posteriori error estimation in finite element methods. *Acta Numerica*, 10:1–102, 2001. doi: 10.1017/S0962492901000010.

J. Frédéric Bonnans and Alexander Shapiro. *Perturbation Analysis of Optimization Problems*. Springer, 2000. doi: 10.1007/978-1-4612-1394-9.

Miranda Boutilier, Konstantin Brenner, and Larissa Miguez. Learning local Dirichlet-to-Neumann maps of nonlinear elliptic PDEs with rough coefficients. *ESAIM: Proceedings and Surveys*, 81: 16–32, 2025. doi: 10.1051/proc/202581016.

Wojciech Marian Czarnecki, Simon Osindero, Max Jaderberg, Grzegorz Świrszcz, and Razvan Pascanu. Sobolev training for neural networks. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://papers.nips.cc/paper/7015-sobolev-training-for-neural-networks.

Peter Deuflhard. *Newton Methods for Nonlinear Problems: Affine Invariance and Adaptive Algorithms*, volume 35 of *Springer Series in Computational Mathematics*. Springer, 2004. ISBN 978-3-540-21099-3. URL https://link.springer.com/book/9783540210993.

Niccolò Discacciati and Jan S. Hesthaven. Model reduction of coupled systems based on non-intrusive approximations of the boundary response maps. *Computer Methods in Applied Mechanics and Engineering*, 420:116770, 2024. doi: 10.1016/j.cma.2024.116770.

Jens L. Eftang and Anthony T. Patera. A port-reduced static condensation reduced basis element method for large component-synthesized structures: approximation and a posteriori error estimation. *Advanced Modeling and Simulation in Engineering Sciences*, 1:3, 2014. doi: 10.1186/2213-7467-1-3. URL https://link.springer.com/article/10.1186/2213-7467-1-3.

Gianluca Fabiani, Hannes Vandecasteele, Somdatta Goswami, Constantinos Siettos, and Ioannis G. Kevrekidis. Enabling local neural operators to perform equation-free system-level analysis. *Nature Machine Intelligence*, 8:1127–1141, 2026. doi: 10.1038/s42256-026-01265-1. URL https://www.nature.com/articles/s42256-026-01265-1.

Tony Haines and Ke Shi. A neural-network-based multiscale hybridizable discontinuous Galerkin method for solving PDEs in porous media, 2026. URL https://arxiv.org/abs/2608.25850v2. Preprint; version 2, September 4, 2026.

Jianing Huang, Kaixuan Zhang, Youjia Wu, and Ze Cheng. Operator learning with domain decomposition for geometry generalization in PDE solving. In *International Conference on Learning Representations*, 2026.

Hongyue Jiang, Jianjiang Zhan, Chenzhuo Zhang, and Fan Wang. Convex neural energy elements: Monolithic finite-element assembly of geometry-parameterized neural operators with stability and error guarantees, 2026. URL https://arxiv.org/abs/2608.02036. Preprint.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to PDEs. *Journal of Machine Learning Research*, 24(89):1–97, 2023. URL https://jmlr.org/papers/v24/21-1524.html.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021. URL https://arxiv.org/abs/2010.08895.

Shuo Ling, Wenjun Ying, and Han Zhou. Principal-part decomposition for neural operator learning of Dirichlet-to-Neumann maps, 2026. URL https://arxiv.org/abs/2606.25952. Preprint.

Samuel Lippl and Kimberly Stachenfeld. When does compositional structure yield compositional generalization? a kernel theory. In *International Conference on Learning Representations*, 2025.

Michael L. Littman, Richard S. Sutton, and Satinder Singh. Predictive representations of state. In *Advances in Neural Information Processing Systems*, volume 14, 2001.

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*, 3:218–229, 2021. doi: 10.1038/s42256-021-00302-5. URL https://www.nature.com/articles/s42256-021-00302-5.

Jelena Luketina, Jack Lanchantin, Sainbayar Sukhbaatar, and Arthur Szlam. Compositional interfaces for compositional generalization. In *Proceedings of the 3rd Conference on Lifelong Learning Agents*, volume 274 of *Proceedings of Machine Learning Research*, pp. 692–709. PMLR, 2025. URL https://proceedings.mlr.press/v274/luketina25a.html.

Per-Gunnar Martinsson. The hierarchical Poincaré–Steklov (HPS) solver for elliptic PDEs: A tutorial, 2015. URL https://arxiv.org/abs/1506.01308.

Nick McGreivy and Ammar Hakim. Weak baselines and reporting biases lead to overoptimism in machine learning for fluid-related partial differential equations. *Nature Machine Intelligence*, 6 (10):1256–1269, 2024. doi: 10.1038/s42256-024-00897-5. URL https://www.nature.com/articles/s42256-024-00897-5.

Weihang Ouyang, Yeonjong Shin, Si-Wei Liu, and Lu Lu. NOEM: efficient and scalable finite element method enabled by reusable neural operators. *Nature Computational Science*, 6(4):417–429, 2026. doi: 10.1038/s43588-026-00974-2. URL https://www.nature.com/articles/s43588-026-00974-2.

Helen Qu, Rudy Morel, Michael McCabe, Alberto Bietti, François Lanusse, Shirley Ho, and Yann LeCun. Representation learning for spatiotemporal physical systems, 2026. URL https://arxiv.org/abs/2603.13227. ICLR 2026 Workshop on AI & PDE.

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

Alexander Rudikov, Vladimir Fanaskov, Sergei Stepanov, Buzheng Shan, Ekaterina Muravleva, Yalchin Efendiev, and Ivan Oseledets. Locally subspace-informed neural operators for efficient multiscale PDE solving. In *International Conference on Learning Representations*, 2026.

Kathrin Smetana and Anthony T. Patera. Optimal local approximation spaces for component-based static condensation procedures. *SIAM Journal on Scientific Computing*, 38(5):A3318–A3356, 2016. doi: 10.1137/15M1009603. URL https://epubs.siam.org/doi/10.1137/15M1009603.

Amy Zhang, Rowan McAllister, Roberto Calandra, Yarin Gal, and Sergey Levine. Learning invariant representations for reinforcement learning without reconstruction. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=-2FCwDKRREu.

SUPPLEMENTARY MATERIAL

The supplement follows the argument of the main paper. Appendix A specifies the cubic and sinh frozen-decoder interventions, the three-by-three model crossing, matched auxiliary supervision and checkpoint selection. The field evaluation reports relative L^2 and energy-norm errors together. Appendix B develops the contextual quotient, exact and approximate sufficiency, risks of reuse, and elliptic elimination. Appendix C develops propagation bounds, checks actual learned errors with reference and consistent evaluator tangents, and gives the controlled protocol and all 72 case results. Appendix D describes the analytic and learned evaluators. Appendix E collects the earlier target, breadth, and port studies. Appendix F treats computational consequences and parent relations. The source package includes the recorded rows and configurations used to generate every new numerical figure and table.

Field errors evaluate the requested output. Interface-trace errors evaluate the composed equilibrium. Relative prediction error evaluates the propagation theory by comparing its predicted displacement vector with the observed one. Spearman correlation measures the ordering of displacement magnitudes, and cosine similarity measures alignment of the displacement vectors. The local, assembled and propagated norms in Appendix C.3 are candidate predictors of the same observed displacement.

Study	Physical units	Question answered
Frozen-output intervention	32 cubic and 16 sinh cases, two paired meshes, three decoders per family	Effect of response substitution with each decoder fixed
Auxiliary supervision and model crossing	Same 32 cubic cases, paired continuations and nine model pairings	Interaction supervision and robustness across trained models
Actual-error diagnostic	Same 32 cubic and 16 sinh cases, three decoders per family, shared relations counted once	Propagation of defects made by trained interfaces
Controlled error geometry	72 existing 1D assemblies, five magnitudes and three directions	Effect of response-error direction with unchanged tangents
Earlier target and breadth studies	Separate 2D catalogs and family-specific trained weights	Parameterization and evaluator choice within a response interface
Port and computational studies	Separate 2D transfer, audit, and query workloads	Interface capacity and execution cost

A FROZEN-OUTPUT INTERVENTION AND HELD-OUT CONFIRMATION

A.1 PHYSICAL PROBLEM AND THE THREE RESPONSE EVALUATORS

The cubic intervention uses $\phi(u) = u^3$ in (1) on squares of width $h = 0.25$. The local parameter vector is $p = (a, c, \gamma, f_0, f_x, f_y, h)$, with reference-square forcing $f = f_0 + f_x \cos(\pi\widehat{x}) + f_y \cos(\pi\widehat{y})$. The assembly is a 4×4 grid with fixed exterior Dirichlet data and shared interior traces. Each scalar component has a 16-dimensional port at both mesh resolutions. The discretizations use 17 or 33 nodal points per local coordinate axis.

Let $r_K(u)$ denote the full local weak residual, and let P_K prolong port coefficients to nodal boundary values. The three response evaluators are

$$\mathcal{T}_K^{\mathrm{field}}(\lambda) = P_K^{\top}[r_K(\mathcal{D}_K(\lambda, p_K))]_E, \tag{10}$$

$$\mathcal{T}_K^{\mathrm{learned}}(\lambda) = \mathcal{T}_K^{[2]}(\lambda) + \omega(s_K) s_K Q_\theta^{\mathrm{phys}}(\lambda, p_K), \tag{11}$$

$$\mathcal{T}_K^{\mathrm{HF}}(\lambda) = P_K^{\top}[r_K(u_K^{\star}(\lambda, p_K))]_E. \tag{12}$$

The field-induced evaluator evaluates the physical boundary residual on the predicted field. It differentiates this complete map with forward-mode Jacobian-vector products for all 16 port columns. It performs no interior re-equilibration. This response is distinct from differentiating an energy

composed with the approximate field predictor. The HF evaluator solves the nonlinear interior equations at the requested trace and returns the condensed tangent. The learned evaluator is specified in Appendix D.

All evaluators enter the same global residual assembly and line-searched Newton solver. The initial free trace is zero. The iteration cap is 18, the normalized residual tolerance is 10^{-9}, and the line search uses ten registered trial steps. Execution proceeds one test case at a time, with local microbatch size two. The learned response is evaluated without Jacobian symmetrization, nonlinear polishing, or local HF fallback. After each global solve, every component field is reconstructed by the identical frozen $\mathcal{D}_K$ at that arm's solved trace.

A.2 The strengthened field reconstructor

The reconstructor has a linear lifting, a forced linear anchor, and a neural interior correction. In natural trace coordinates $z = R_K^{1/2}\lambda$ it returns

$$\mathcal{D}_K(z, p; x, y) = u_{f,K}(x, y) + (B_K z)(x, y) + 16x(1-x)y(1-y)N_\phi(z, p, x, y). \tag{13}$$

Coordinates are on $[0, 1]^2$. The bubble term vanishes on the boundary, so the prescribed trace is exact. The network concatenates normalized trace and physical context features with spatial coordinates and sine/cosine features at frequencies $\{1, 2, 4\}\pi$. It has width 256, four residual blocks, SiLU activations, and a scalar output head. A residual block applies two linear transformations and SiLU activations, with the residual sum divided by $\sqrt{2}$. Computation uses double precision.

Define the positive quadratic field energy

$$\mathcal{E}_K(v) = a_K v^\top K_K v + h_K^2 c_K v^\top M_K v. \tag{14}$$

If $L_{K,I} = C_K C_K^\top$ is the Cholesky factorization of the linear interior operator and $r_{K,I}$ is the interior residual, the training objective averages

$$\frac{\mathcal{E}_K(\mathcal{D}_K - u_K^{\mathrm{HF}})}{\mathcal{E}_K(u_K^{\mathrm{HF}}) + \tau_E} + 0.1\frac{\left\|C_K^{-1} r_{K,I}(\mathcal{D}_K)\right\|_2^2}{\mathcal{E}_K(u_K^{\mathrm{HF}}) + \tau_E}. \tag{15}$$

The fixed energy floor comes from local training targets. Neither term uses boundary response labels or response-derived target scales. The interior term is a physical residual penalty, so the field training contains more than value regression alone.

The final D model extends a base checkpoint selected at step 7,750 of an 8,000-update local training run. The extension preserves the architecture, normalization, loss, and starting optimizer state. It adds 576 independently sampled local parameter configurations per mesh to the original 192. Four probes per configuration give 3,072 training queries per mesh. New configurations are balanced over a 4×4 grid of training-only quartiles in $\gamma h^2/a$ and physical forcing magnitude. The validation and test pools remain fixed. Training then runs 10,000 additional updates with a cosine learning-rate schedule from 3×10^{-4} to 3×10^{-6}, microbatch two, accumulation four, weight decay 10^{-6}, and gradient clipping at two.

D is the fixed final iterate of this extension, with seed 2026091031. Its mean local validation field errors are 1.132875% at 17 nodes and 0.848926% at 33 nodes. The subsequent frozen-output intervention explicitly uses this final model as its fixed reconstructor. No response-trained continuation or latent readout is used. The learned-relation evaluator is the existing cubic seed 2026090617 model with JVP weight 0.25. The models have different capacities and training histories.

A.3 Development, freezing, and independent confirmation

A development replay first evaluated 32 test cases with the same three response evaluators. Its mean field errors were 3.450635%, 1.878402%, and 1.878430% for the field-induced, learned-relation, and HF responses. All 32 development cases improved. These observations informed the final intervention protocol.

Before opening the new confirmation cohort, the two trained models, activation, solver, aggregation, and three-arm comparison were fixed. Confirmation uses the previously generated 32-case catalog

with seed 2026091039 and replicate offset 30000. It contains eight cases per pattern and has no complete parameter-tensor overlap with development. No outcome from this cohort selects a model or a numerical setting. This initial confirmation uses one fixed pair of models. The subsequent replication below trains two additional decoders on the same local training data. The 32 test cases are the evaluation units, with two mesh resolutions paired within each case.

The recorded directional criterion required lower field error on at least 28 of 32 cases and positive mean paired log improvement. The recorded structural comparison required relative proximity to the HF-response reconstruction error within 10% on at least 28 cases. These are preconfirmation criteria for this intervention. The observed counts are 32 for both, and the maximum relative proximity discrepancy is 0.611%. The numerical errors and paired outcomes below provide the primary evidence.

A.4 Reference, statistical definitions, and complete results

The reference trace is the HF equilibrium of the same mesh and the same 16-dimensional local port. Its reference field $u^\star$ uses nonlinear HF interior reconstruction. For a candidate u, the relative energy-norm field error is

$$e_E = \left(\frac{\sum_K \mathcal{E}_K(u_K - u_K^\star)}{\max(\sum_K \mathcal{E}_K(u_K^\star), 10^{-20})} \right)^{1/2}. \tag{16}$$

The relative interface-trace error uses the assembled positive linear metric with the recorded denominator floor,

$$e_\Lambda = \frac{\left\| \widehat{\Lambda} - \Lambda^\star \right\|_R}{\max(\|\Lambda^\star\|_R, 0.05)}.$$

The HF-response arm still reconstructs with $\mathcal{D}$, so its field error is generally positive. The comparison uses a same-port discrete reference throughout.

Let $e_{p,n}^{(a)}$ be the error of arm a on case p and resolution n. The case value is $(e_{p,17}^{(a)} + e_{p,33}^{(a)})/2$. Means, medians, and quantiles are computed from the 32 case values. Quantiles use linear interpolation. The headline reduction is $1 - \mathrm{mean}_p\, e_p^{(B)} / \mathrm{mean}_p\, e_p^{(A)}$. It differs from the mean of individual percentage reductions. The mean paired log ratio is 0.595414 and the geometric ratio is 1.813781. The relative B/C discrepancy is $|e_p^{(B)} - e_p^{(C)}| / \max(e_p^{(C)}, 10^{-12})$.

Table 3: Confirmation relative energy-norm errors (%) over the 32 case averages.

Interface	Mean	Median	P90	P95	Max
Field-induced	3.601529	2.861372	7.848452	8.364175	11.034613
Learned relation	1.974668	1.746110	4.023371	4.711851	6.096607
HF response	1.974967	1.745653	4.022760	4.710426	6.095875

Table 4: The effect is present in every registered layout stratum. Energy-norm errors are percentages and the final column is reduction of the stratum mean. Each stratum has eight improved cases.

Pattern	Cases	Field-induced	Learned	HF	Reduction (%)
checkerboard	8	3.1817	1.8769	1.8768	41.01
inclusion	8	3.8574	1.8884	1.8906	51.05
repeated block	8	3.3392	1.9469	1.9467	41.69
stripes	8	4.0279	2.1865	2.1858	45.72

All 192 candidate solves converge, comprising 64 per arm. The 64 distinct HF references converge as well. There are 64 equal-input field-repeat checks, with zero maximum difference, and the trained field state is unchanged across evaluation. The confirmation runs in FP64 with PyTorch 2.11.0 and CUDA 12.8 on the shared RTX 5090 server, using one GPU and two CPU threads. The runtime is 24.655 minutes and peak allocated tensor memory is 1.266 GiB. These are execution records for the intervention, without a throughput claim.

Table 5: All 32 confirmation cases. Relative energy-norm errors are percentages after averaging the two meshes. Reduction is computed separately for each case.

Test case	Field-induced	Learned	HF	Reduction (%)
Checker 01	1.6148	0.6966	0.6966	56.86
Checker 02	4.4631	2.4219	2.4217	45.73
Checker 03	0.6365	0.5346	0.5346	16.01
Checker 04	2.9994	1.3256	1.3256	55.80
Checker 05	3.1654	1.9286	1.9291	39.07
Checker 06	8.0147	4.8590	4.8571	39.37
Checker 07	1.6585	1.0019	1.0019	39.59
Checker 08	2.9007	2.2467	2.2477	22.55
Inclusion 01	3.8499	1.7504	1.7498	54.53
Inclusion 02	6.2799	3.3202	3.3179	47.13
Inclusion 03	3.3351	2.1050	2.1136	36.88
Inclusion 04	5.1010	1.6782	1.6783	67.10
Inclusion 05	0.7589	0.4406	0.4406	41.94
Inclusion 06	2.5655	1.8148	1.8260	29.26
Inclusion 07	5.3040	2.3457	2.3470	55.78
Inclusion 08	3.6645	1.6520	1.6515	54.92
Repeated 01	2.2798	1.9135	1.9137	16.07
Repeated 02	2.3226	1.3494	1.3495	41.90
Repeated 03	2.4091	1.1047	1.1047	54.14
Repeated 04	1.0136	0.5466	0.5466	46.08
Repeated 05	2.7441	2.3097	2.3087	15.83
Repeated 06	8.3590	4.5915	4.5904	45.07
Repeated 07	1.2332	0.5644	0.5644	54.23
Repeated 08	6.3520	3.1956	3.1952	49.69
Stripe 01	8.3705	4.0966	4.0964	51.06
Stripe 02	0.4671	0.2891	0.2891	38.11
Stripe 03	11.0346	6.0966	6.0959	44.75
Stripe 04	2.8220	1.7418	1.7415	38.28
Stripe 05	0.7457	0.3600	0.3600	51.73
Stripe 06	2.2576	0.8178	0.8178	63.78
Stripe 07	5.2882	3.3639	3.3597	36.39
Stripe 08	1.2376	0.7263	0.7263	41.31

A.5 Independent decoder replication

Seeds 2026091131 and 2026091231 each start from independent weights and an independent AdamW optimizer. Each completes the original 8,000-step base training and continues its own selected base state and optimizer for all 10,000 extension steps. The final iterate is frozen unconditionally. Both base checkpoints are selected at step 7,750. Architecture, losses, local training data, and validation rules match the original decoder. The relation evaluator remains seed 2026090617 in all repetitions.

Each frozen decoder reconstructs the field in all three response arms on the same 32 test cases and two meshes. This is training-seed replication on the original held-out cohort. Every seed improves all 32 cases, with reductions of mean field error of 45.17%, 43.12%, and 46.45%. All 576 candidate evaluations converge. Table 6 reports each seed separately. Averaging seeds within each test case gives field errors of 3.777176%, 2.081224%, and 2.080751%, and a 44.90% reduction of the first two means. The 96 seed–case pairs share 32 physical inputs.

The replication uses the same FP64 solver and hardware as the original confirmation. The complete per-seed and per-case measurements follow.

Table 6: Three independent field decoders under the same interface intervention. Relative energy-norm errors are percentages after mesh averaging within each of 32 cases. The B/C gap is the relative difference between learned-response and HF-response field errors. Every decoder is evaluated with the same learned relation.

Decoder seed	Field-induced	Learned	HF	Reduction (%)	Improved	Max B/C gap (%)
2026091031	3.601529	1.974668	1.974967	45.1714	32/32	0.6106
2026091131	3.889843	2.212424	2.210745	43.1230	32/32	1.1657
2026091231	3.840157	2.056579	2.056540	46.4454	32/32	0.2903

Table 7: Independent decoder training and final local validation energy-norm errors. The two new seeds use their own base and continuation optimizer states.

Seed	Base step	Final step	17-node error (%)	33-node error (%)
2026091031	7750	10000	1.132875	0.848926
2026091131	7750	10000	1.234826	0.763914
2026091231	8000	10000	1.187559	0.852771

Table 8: Relative energy-norm errors for all three decoder seeds on the same 32 test cases. Each entry averages two meshes. Seed suffixes identify 2026091031, 2026091131, and 2026091231.

Seed suffix	Test case	Field-induced (%)	Learned (%)	HF (%)
1031	Repeated 01	2.2798	1.9135	1.9137
1031	Stripe 01	8.3705	4.0966	4.0964
1031	Checker 01	1.6148	0.6966	0.6966
1031	Inclusion 01	3.8499	1.7504	1.7498
1031	Stripe 02	0.4671	0.2891	0.2891
1031	Inclusion 02	6.2799	3.3202	3.3179
1031	Repeated 02	2.3226	1.3494	1.3495
1031	Stripe 03	11.0346	6.0966	6.0959
1031	Checker 02	4.4631	2.4219	2.4217
1031	Checker 03	0.6365	0.5346	0.5346
1031	Inclusion 03	3.3351	2.1050	2.1136
1031	Repeated 03	2.4091	1.1047	1.1047
1031	Inclusion 04	5.1010	1.6782	1.6783
1031	Stripe 04	2.8220	1.7418	1.7415
1031	Inclusion 05	0.7589	0.4406	0.4406
1031	Stripe 05	0.7457	0.3600	0.3600
1031	Repeated 04	1.0136	0.5466	0.5466
1031	Checker 04	2.9994	1.3256	1.3256
1031	Stripe 06	2.2576	0.8178	0.8178
1031	Checker 05	3.1654	1.9286	1.9291
1031	Stripe 07	5.2882	3.3639	3.3597
1031	Repeated 05	2.7441	2.3097	2.3087
1031	Inclusion 06	2.5655	1.8148	1.8260
1031	Stripe 08	1.2376	0.7263	0.7263
1031	Inclusion 07	5.3040	2.3457	2.3470
1031	Repeated 06	8.3590	4.5915	4.5904
1031	Inclusion 08	3.6645	1.6520	1.6515
1031	Repeated 07	1.2332	0.5644	0.5644
1031	Repeated 08	6.3520	3.1956	3.1952
1031	Checker 06	8.0147	4.8590	4.8571
1031	Checker 07	1.6585	1.0019	1.0019
1031	Checker 08	2.9007	2.2467	2.2477
1131	Repeated 01	2.0375	1.6955	1.6941
1131	Stripe 01	8.8412	4.6752	4.6749
1131	Checker 01	1.0869	0.6222	0.6222
1131	Inclusion 01	6.6825	3.2370	3.2361
1131	Stripe 02	0.8481	0.3503	0.3503

Seed suffix	Test case	Field-induced (%)	Learned (%)	HF (%)
1131	Inclusion 02	9.5566	5.7181	5.7149
1131	Repeated 02	1.5468	1.0530	1.0531
1131	Stripe 03	11.5900	6.9301	6.9293
1131	Checker 02	5.9132	3.0150	3.0146
1131	Checker 03	1.2163	0.5414	0.5414
1131	Inclusion 03	1.7714	1.2215	1.2134
1131	Repeated 03	3.1349	1.2618	1.2618
1131	Inclusion 04	3.2832	1.2026	1.2026
1131	Stripe 04	2.9626	2.0211	2.0205
1131	Inclusion 05	1.1575	0.3626	0.3626
1131	Stripe 05	0.6570	0.4075	0.4075
1131	Repeated 04	1.6083	0.6062	0.6062
1131	Checker 04	2.0333	1.0221	1.0221
1131	Stripe 06	2.6626	0.9424	0.9424
1131	Checker 05	1.7839	1.3337	1.3340
1131	Stripe 07	5.6601	4.0372	4.0322
1131	Repeated 05	3.0311	2.3549	2.3500
1131	Inclusion 06	2.8703	2.0368	2.0133
1131	Stripe 08	1.1939	0.7650	0.7650
1131	Inclusion 07	2.4962	1.2480	1.2495
1131	Repeated 06	9.6196	5.6297	5.6283
1131	Inclusion 08	6.4501	3.0650	3.0642
1131	Repeated 07	0.8835	0.5356	0.5356
1131	Repeated 08	7.4457	3.8895	3.8890
1131	Checker 06	9.9138	6.2234	6.2208
1131	Checker 07	2.5816	1.1039	1.1039
1131	Checker 08	1.9551	1.6893	1.6880
1231	Repeated 01	2.1162	1.8466	1.8461
1231	Stripe 01	9.2534	4.4338	4.4336
1231	Checker 01	1.2602	0.6009	0.6009
1231	Inclusion 01	6.3606	2.5597	2.5591
1231	Stripe 02	0.5452	0.3477	0.3477
1231	Inclusion 02	7.9508	3.8406	3.8385
1231	Repeated 02	1.2294	1.1148	1.1148
1231	Stripe 03	11.5238	6.4788	6.4781
1231	Checker 02	5.8788	2.7262	2.7260
1231	Checker 03	1.2137	0.5664	0.5664
1231	Inclusion 03	2.7348	1.8307	1.8359
1231	Repeated 03	3.8647	1.4435	1.4435
1231	Inclusion 04	2.5672	1.1104	1.1105
1231	Stripe 04	3.4879	2.0516	2.0512
1231	Inclusion 05	0.7147	0.3127	0.3127
1231	Stripe 05	0.8361	0.4031	0.4031
1231	Repeated 04	1.8324	0.6781	0.6781
1231	Checker 04	1.4568	0.9996	0.9996
1231	Stripe 06	3.4147	1.1401	1.1401
1231	Checker 05	1.7158	1.5314	1.5316
1231	Stripe 07	5.7805	3.9301	3.9269
1231	Repeated 05	2.7403	2.3215	2.3196
1231	Inclusion 06	2.7264	1.9107	1.9162
1231	Stripe 08	1.5204	0.8319	0.8319
1231	Inclusion 07	2.6944	1.5262	1.5274
1231	Repeated 06	9.0189	4.7569	4.7559
1231	Inclusion 08	6.2243	2.4689	2.4684
1231	Repeated 07	1.0865	0.5055	0.5055
1231	Repeated 08	7.4626	3.4895	3.4891
1231	Checker 06	8.5231	4.8667	4.8649
1231	Checker 07	3.1390	1.2655	1.2655
1231	Checker 08	2.0113	1.9201	1.9203

A.6 THE FROZEN-DECODER INTERVENTION IN SINH

We repeat the intervention for $-\nabla \cdot (a\nabla u) + cu + \gamma(\sinh u - u) = f$. The local response uses the full nonpolynomial reaction. The components have width $h = 0.25$, Q1 meshes with 17 or 33 nodes per axis and a 16-dimensional port. The existing 16-case catalog contains checkerboard, inclusion, repeated-block and stripe layouts in 4×4 assemblies.

Three independent field models use seeds 2026091251, 2026091252 and 2026091253. The width-256 decoder has four residual blocks, coordinate frequencies $1, 2, 4$, a linear lifting and a boundary-vanishing bubble factor. Its isolated queries reproduce the original sinh sampling manifest, with 819 training and 205 validation records across the two meshes. HF interior fields are computed for these queries. The field objective includes an interior residual penalty of weight 0.1. Response labels are excluded from field training.

Training runs for 18,000 updates, using microbatches of two with four-step accumulation, weight decay 10^{-6} and gradient clipping at two. The learning rate decreases from 3×10^{-4} to 3×10^{-6}. Validation occurs every 500 updates. Each model selects the checkpoint with the smallest worst-resolution mean validation field error, with the earliest checkpoint breaking a tie. All three models freeze before global evaluation. Their final selected local-validation scores are 0.1573%, 0.1589% and 0.1777%.

The dedicated relation is the original Analytic-2+NN evaluator with seed 2026090837. Its 20–128–128–16 network was trained for 1,500 updates with batch size 64. It retains the response formula and trace-radius fallback of the earlier sinh study in Appendix E.2. These field and relation models have different capacities and training budgets. The intervention compares their realized responses while fixing the downstream numerical system.

For each field model, the field-induced, learned-relation and HF responses use the same Newton solver, zero initialization, 25 updates, residual tolerance 10^{-9} and 12 line-search halvings. Tangents retain their full nonsymmetric form. Each arm reconstructs through the same frozen decoder. The reference is the nonlinear same-mesh, same-port HF solution, and field errors use the energy norm of Appendix A. The HF-response arm therefore measures the decoder at the reference trace.

Table 9: Sinh response substitution with each output decoder fixed. Entries average the two meshes within each of 16 cases before taking the case mean. Field columns report relative energy-norm errors. The relation evaluator is shared by all three decoders.

Decoder seed	Response evaluator	Trace (%)	Energy (%)
2026091251	Field-induced	0.1862	0.2040
2026091251	Analytic-2+NN	1.4497	1.4047
2026091251	HF	0.0000	0.0940
2026091252	Field-induced	0.1983	0.2163
2026091252	Analytic-2+NN	1.4497	1.4063
2026091252	HF	0.0000	0.0997
2026091253	Field-induced	0.1848	0.2039
2026091253	Analytic-2+NN	1.4497	1.4055
2026091253	HF	0.0000	0.0974

HF responses improve all 16 cases for each decoder. The three-seed mean field error falls from 0.208060% to 0.097021%, a 53.37% reduction. The current learned relation increases it to 1.405507%. The corresponding interface means are 0.189734% for the field-induced response and 1.449679% for the learned relation. The latter gives larger interface and field errors on every case for each seed. Its geometric mean A/B error ratios are 0.123101 [0.096053, 0.162813] for interface error and 0.143625 [0.115059, 0.184025] for field error, where A is field-induced and B is the learned relation.

All 288 candidate evaluations converge, and all equal-input decoder repeats agree. The learned evaluator makes zero HF fallbacks in 13,056 response calls. Its maximum queried trace radius is 3.7941, below the fallback threshold of five. The reference and response computations retain the full sinh reaction. These results distinguish the accuracy of a numerical evaluator from the external information required for composition.

A.7 RELATIVE L^2 AND ENERGY-NORM FIELD ERRORS

We evaluate the same frozen decoders at the saved equilibria of all three response arms. This supplements the intervention with the standard relative L^2 error

$$e_{L^2} = \frac{\|u - u^\star\|_{L^2(\Omega)}}{\|u^\star\|_{L^2(\Omega)}} = \left(\frac{\sum_K h_K^2 (u_K - u_K^\star)^\top M_K (u_K - u_K^\star)}{\sum_K h_K^2 (u_K^\star)^\top M_K u_K^\star} \right)^{1/2}. \tag{17}$$

Here M_K is the Q1 mass matrix on the reference square. The factor h_K^2 maps its integral to the physical component. The reference is the nonlinear HF field at the same-mesh, same-port HF equilibrium. The integral covers the full assembled domain. This norm measures solution-value error, while the energy norm in (16) also weights gradient error through $a_K \|\nabla v\|_2^2 + c_K v^2$.

Table 10: **HF responses improve both field norms.** Values are mean ± sample SD of three decoder-seed means, in percent. Each case first averages nodes17/33; cases then receive equal weight within each seed. Cubic has 32 cases and sinh 16, shared across the three seeds. Each family uses one shared learned-relation evaluator.

PDE	Response evaluator	Relative L^2 (%)	Relative energy (%)
Cubic	Field-induced	3.4914 ± 0.1995	3.7772 ± 0.1541
Cubic	Learned relation	0.5349 ± 0.0251	2.0812 ± 0.1208
Cubic	HF response	0.5365 ± 0.0267	2.0808 ± 0.1197
Sinh	Field-induced	0.1600 ± 0.0092	0.2081 ± 0.0071
Sinh	Learned relation	1.3658 ± 0.0010	1.4055 ± 0.0008
Sinh	HF response	0.0206 ± 0.0012	0.0970 ± 0.0029

HF responses reduce mean relative L^2 error by 84.6% in cubic and 87.1% in sinh. Every case improves for each decoder, giving 96 of 96 and 48 of 48 paired seed–case comparisons. The learned relation also improves all cubic comparisons and increases error in every sinh comparison. The two norms therefore agree on the direction of response-substitution effects. Their different magnitudes reflect the additional sensitivity of the energy norm to gradients in the reconstructed field.

The supplement contains 864 mesh-level evaluations from the original checkpoints and final interface states. Reconstructing the local fields requires no training, checkpoint selection, or additional global Newton solves. Mass-matrix and quadrature evaluations of the squared L^2 error agree within 4.3×10^{-16} relative difference. Replayed energy and trace errors agree with the original measurements within 1.4×10^{-17} and 1.6×10^{-15} absolute error, respectively. The field snapshots and per-mesh, per-case and per-seed records accompany the evaluation data.

A.8 MATCHED RESPONSE AND TANGENT SUPERVISION FOR THE FIELD MODEL

Each original cubic decoder starts two 3,000-update continuations. The two arms share the starting weights, architecture, sampled training batches and optimizer schedule. One retains the field objective. The other adds response and JVP losses to that objective, with responses induced by its predicted field. The added supervision therefore changes the field model's interaction behavior through the same field-to-response computation used during evaluation.

For each auxiliary loss, its weight is half the median ratio of the base-objective gradient norm to the auxiliary gradient norm. Calibration uses the first decoder seed and 16 microbatches per mesh with sampling seed 2026091241. The resulting response and JVP weights are 0.119344 and 0.010571, shared unchanged across all three seeds. The learning rate decreases from 3×10^{-5} to 3×10^{-6}, with microbatch two and accumulation four. Both arms retain the final 3,000-update checkpoint unconditionally. The source package records the calibration gradients, weights and per-seed local diagnostics.

Global evaluation retains the 32 cubic cases, two meshes, original Newton settings and frozen reference responses. To distinguish a change in interaction from a change in the output model, each continuation is evaluated both through its own decoder and through the original frozen decoder at

Table 11: Relative field errors (%) for each decoder seed. Each value is the mean over the same mesh-paired cases as Table 10.

PDE	Decoder seed	Response evaluator	L^2 (%)	Energy (%)
Cubic	2026091031	Field-induced	3.277554	3.601529
Cubic	2026091031	Learned relation	0.510021	1.974668
Cubic	2026091031	HF response	0.510162	1.974967
Cubic	2026091131	Field-induced	3.672582	3.889843
Cubic	2026091131	Learned relation	0.560288	2.212424
Cubic	2026091131	HF response	0.563613	2.210745
Cubic	2026091231	Field-induced	3.523930	3.840157
Cubic	2026091231	Learned relation	0.534424	2.056579
Cubic	2026091231	HF response	0.535626	2.056540
Sinh	2026091251	Field-induced	0.159732	0.204041
Sinh	2026091251	Learned relation	1.366856	1.404688
Sinh	2026091251	HF response	0.019340	0.093957
Sinh	2026091252	Field-induced	0.169254	0.216257
Sinh	2026091252	Learned relation	1.364806	1.406290
Sinh	2026091252	HF response	0.021740	0.099667
Sinh	2026091253	Field-induced	0.150898	0.203883
Sinh	2026091253	Learned relation	1.365877	1.405543
Sinh	2026091253	HF response	0.020777	0.097439

its equilibrium. Let $\bar{e}^{(a)}_{p,s}$ average the two meshes for case p, seed s and arm a. For each case, form

$$\ell_p = \frac{1}{3}\sum_s \log \frac{\bar{e}^{(\mathrm{field})}_{p,s}}{\bar{e}^{(\mathrm{auxiliary})}_{p,s}}.$$

The geometric mean error ratio is $\exp(\mathrm{mean}_p\, \ell_p)$. Its 95% percentile interval uses 10,000 resamples of the 32 test cases with seed 2026091243. Seeds and meshes remain paired within each resampled case.

Table 12: Matched field-model continuation. Arithmetic error means and paired geometric mean error ratios are distinct summaries. A ratio above one favors response/JVP supervision. Intervals resample test cases.

Metric	Field only (%)	With auxiliary (%)	Error ratio	95% interval
Interface trace	3.4783	3.2046	1.024	[0.940, 1.115]
Energy, fixed original $\mathcal{D}$	3.7196	3.5134	1.016	[0.962, 1.073]
Energy, continued $\mathcal{D}$	3.7138	3.4332	1.040	[0.980, 1.105]

The primary interface interval includes one, with 16 of 32 cases improving after averaging the seed log gains. The common-decoder and continued-decoder field intervals also include one. All 384 solves converge and all 384 common-decoder repeats agree. The largest ratio of local field-error P95s between the two arms is 1.1232. Under this continuation design, adding response and tangent supervision leaves substantial composition error and gives no stable paired gain. The data describe this objective and budget.

A.9 Crossing decoder and relation checkpoints

The three cubic decoders are crossed with the existing relation checkpoints from seeds 2026090617, 2026090717 and 2026090817. Each pairing replaces the decoder's field-induced response and retains that decoder for reconstruction. The relation equilibrium is reused across decoders. The checkpoints retain their historical selected training policies, including their auxiliary-loss weights, so the experiment tests robustness across these nine model pairings.

Mean field error decreases from 3.777176% to 2.081222%. Averaging all nine log error ratios within each case gives a geometric mean error ratio of 1.823634 with interval [1.682580, 1.968650]. The bootstrap uses 10,000 case resamples with seed 2026091260. All 576 reconstruction records are valid. The inferential sample size remains 32 test cases.

Table 13: Nine cubic decoder/relation pairings. Every pairing improves all 32 cases. Relative energy-norm errors are means in percent. Each error ratio compares the paired decoder's field-induced response with the selected learned response.

Decoder seed	Relation seed	Energy (%)	Error ratio	95% interval
2026091031	2026090617	1.9747	1.8138	[1.678, 1.956]
2026091031	2026090717	1.9747	1.8138	[1.678, 1.956]
2026091031	2026090817	1.9748	1.8136	[1.678, 1.956]
2026091131	2026090617	2.2124	1.8190	[1.664, 1.989]
2026091131	2026090717	2.2126	1.8190	[1.664, 1.989]
2026091131	2026090817	2.2121	1.8194	[1.664, 1.989]
2026091231	2026090617	2.0566	1.8382	[1.660, 2.019]
2026091231	2026090717	2.0566	1.8381	[1.660, 2.018]
2026091231	2026090817	2.0566	1.8381	[1.659, 2.018]

A.10 Local validation and composed-error checkpoint rankings

For each cubic decoder, the candidate set contains its original checkpoint and the field-only continuation at 1,000, 2,000 and 3,000 updates. Local score is the larger of the two mesh-specific mean validation field errors. Composed score averages the two meshes within each of the 32 cases, then averages cases. The checkpoint with the smallest composed score is a retrospective comparator. Global outcomes do not enter training or checkpoint selection.

Table 14: Checkpoint ranking within each training trajectory. The two choice columns show continuation updates or the original checkpoint. Excess error is the composed energy-norm error difference from the retrospectively best candidate in percentage points. Spearman correlation compares the four local and global scores.

Seed	ρ_S	Reversed pairs	Local choice	Global best	Excess error (pp)
2026091031	-0.6	4/6	3000	Original	0.0388
2026091131	+0.8	1/6	3000	3000	0.0000
2026091231	-0.2	3/6	3000	1000	0.0880

Every trajectory contains a reversed candidate pair. Local-score selection differs from the composed-error optimum in two seeds, with excess errors of 0.0388 and 0.0880 percentage points. The pooled correlation across all 12 candidates is positive, 0.6923. Thus the result concerns imperfect selection with modest observed losses. All candidates converge on both meshes for all 32 cases. Checkpoints within a trajectory remain dependent observations.

The follow-ups use FP64 on one NVIDIA RTX 5090 with PyTorch 2.11.0, CUDA 12.8 and two CPU threads. Recorded protocols, all per-case rows and deterministic aggregation code accompany the manuscript source.

B Contextual representations and exact elimination

B.1 Contexts, observables, and replacement

A context C specifies how a component is connected to its surroundings, the external inputs, and the solution convention. An observable O specifies what is measured after composition. Fix a class $\mathcal{X}$ of components with a common interface type and a common set $\mathcal{C}$ of admissible contexts. The deterministic value $O(C[X])$ must be defined for every $(C, X) \in \mathcal{C} \times \mathcal{X}$. Solution-branch selection is part of this convention. Contextual equivalence in (2) is an equivalence relation. Reflexivity, symmetry, and transitivity follow from equality of the observable in each fixed context.

B.2 THE CONTEXTUAL QUOTIENT

The quotient identifies components with the same observable behavior in every admissible context,

$$\pi_{\mathcal{C},O} : \mathcal{X} \longrightarrow \mathcal{X}/\!\sim_{\mathcal{C},O}, \qquad \pi_{\mathcal{C},O}(X) = [X]_{\mathcal{C},O}. \tag{18}$$

This is the standard quotient construction applied to the chosen contextual equivalence. It gives a canonical informational object against which to compare candidate representations.

Proposition 3 (Factorization through a sufficient representation)**.** *For a representation $r : \mathcal{X} \to \mathcal{Z}$, composition sufficiency is equivalent to the existence of a unique map $h : r(\mathcal{X}) \to \mathcal{X}/\!\sim_{\mathcal{C},O}$ satisfying*

$$\pi_{\mathcal{C},O} = h \circ r. \tag{19}$$

It is also equivalent to the existence, for every $C \in \mathcal{C}$, of a map $\widetilde{O}_C$ on $r(\mathcal{X})$ with $O(C[X]) = \widetilde{O}_C(r(X))$.

Proof. If r is sufficient, set $h(r(X)) = [X]_{\mathcal{C},O}$. Equal represented values imply contextual equivalence, so this assignment is well-defined. Every value in $r(\mathcal{X})$ has a preimage, which also proves uniqueness. Conversely, the factorization maps equal represented values to equal contextual classes, proving sufficiency. For each context, set $\widetilde{O}_C(r(X)) = O(C[X])$. The same argument makes this well-defined. Conversely, factorization of every contextual observable through r implies (2) whenever represented values agree. □

Definition 2 (Exact contextual representation)**.** A representation is exact relative to $(\mathcal{C}, O)$ when

$$r(X_1) = r(X_2) \quad \Longleftrightarrow \quad X_1 \sim_{\mathcal{C},O} X_2. \tag{20}$$

For such a representation, h in Proposition 3 is a bijection. Indeed, it is always surjective, and exactness makes it injective. Thus exact representations are recodings of the same contextual quotient. A sufficient representation can retain additional distinctions within one contextual class. For example, if $X = (a, b)$ and every context observes only a, then (a, b) is sufficient and a is exact contextual information.

The quotient is coarsest in the information preorder where r_1 contains the information of r_2 when $r_2 = h \circ r_1$. This is a statement about distinguishable components. Coordinate dimension, computational cost, and statistical learnability require further structure. The construction alone supplies no efficient algorithm for finding the quotient or evaluating its observables. Proposition 1 establishes sufficiency of the boundary response on a fixed trace domain. Identifying it with the quotient would additionally require contexts and observables rich enough to distinguish any two unequal response maps. The finite-port experiments assess capacity relative to their specified references.

B.3 REPRESENTATION, EVALUATOR, COMPOSER, AND DECODER

An executable realization gives the representation a query interface. Let $z = r(X)$ be the represented object and $e(z, \xi)$ its evaluator at query ξ. A composer S_C uses these evaluations to obtain a context state s_C. A decoder D_X produces the requested observable from that state,

$$X \xrightarrow{r} z, \qquad s_C = S_C[e(z, \cdot)], \qquad \widehat{O}_C(X) = D_X(s_C). \tag{21}$$

The middle expression denotes a computation with access to the evaluator. It can issue adaptive queries or solve an implicit equilibrium. This operational structure is additional to the set-theoretic factorization in Proposition 3.

For our elliptic realization, $r(X)$ is the boundary-response map, ξ is a trace, and e supplies a response and its tangent. Analytic, neural, and HF algorithms evaluate the same type of object. The composer enforces compatibility and balance. For an exterior observable, its readout can be part of the context. An interior field requires the additional decoder D_X. If that field is included among the observables used to define equivalence, the reusable payload must contain enough information for both interaction and reconstruction, for example $(r(X), D_X)$.

The frozen-output experiment fixes D_X and S_C while changing e. The controlled geometry experiment changes the evaluator's response error, with the induced state sensitivity described by M. The target screen changes the parameterization used to approximate one response interface. The port study changes the trace space available for communication. These interventions address different parts of (21).

B.4 Approximate sufficiency and risks of reuse

Fix a reference component X and a complete surrogate implementation (r, e, D_X). With a metric d_O on outputs, write

$$\ell_C = d_O\big(\widehat{O}_C(X), O(C[X])\big), \qquad \mathcal{R}_{\text{reuse}}(r, e, D_X; \mu) = \mathbb{E}_{C\sim\mu}\ell_C. \tag{22}$$

Here μ is a specified distribution over admissible contexts, the loss is measurable, and the expectation is assumed to exist. One can also average over a declared joint distribution of components and contexts. Writing the evaluator and decoder explicitly records that a representation alone does not determine approximate execution error. The surrogate is distributionally accurate to tolerance τ when $\mathcal{R}_{\text{reuse}} \le \tau$.

Uniform approximate replacement instead requires

$$\sup_{C\in\mathcal{C}} \ell_C \le \tau.$$

This implies the distributional bound for any probability measure on that context set. Mean risk can be small while rare contexts have large errors. Approximate closeness at a fixed positive tolerance also need not be transitive, so it does not automatically define a quotient like (18).

Two local risks answer different questions. For a prescribed distribution ν of local queries, let $u^\star_X(\xi)$ be the reference output and $a^\star_X(\xi)$ the reference interface response. Define

$$\begin{aligned}\mathcal{R}_{\text{output}}(D_X; \nu) &= \mathbb{E}_{\xi\sim\nu}\, d_U\big(D_X(\xi), u^\star_X(\xi)\big),\\ \mathcal{R}_{\text{interface}}(r, e; \nu) &= \mathbb{E}_{\xi\sim\nu}\, d_T\big(e(r(X), \xi), a^\star_X(\xi)\big).\end{aligned} \tag{23}$$

The losses d_U, d_T and their sampling distributions must be declared. Output risk tests the local prediction, interface risk tests the supplied interaction, and reuse risk follows their effect through the surrounding computation. During composition, the query distribution can itself depend on the evaluator. Even at a fixed query, equal interface-error norms can have different effects through M.

The reported case means are empirical reuse errors on finite recorded cohorts. The two meshes of a test case are averaged before cases receive equal weight. The field-decoder validation scores use a separate local-query distribution. The controlled error study fixes local norms and directions at reference equilibria and checks their consequences through nonlinear re-solves. Its designed directions are probes of the propagation mechanism. They supply no estimate of how frequently trained evaluators produce those directions in a population of contexts.

B.5 Proof of sufficiency of the boundary response

Write the component equations as

$$F_{K,I}(\lambda_K, w_K) = 0, \qquad t_K = F_{K,E}(\lambda_K, w_K).$$

The discretization, trace coordinate adapters, and exterior sign conventions are fixed. Suppose the first equation has a unique smooth solution $w^\star_K(\lambda_K)$ on an admissible domain, and its derivative with respect to w_K is invertible there. This is the local interior-elimination condition in Proposition 1.

Take any full assembly equilibrium on the branch. Its interior equations imply $w_K = w^\star_K(\lambda_K)$. Compatibility expresses the traces as $\lambda_K = E_K\Lambda + \lambda_{K,D}$. Substituting into the interface balance equations gives precisely $G(\Lambda) = 0$ in (4). Conversely, take a root of that reduced system. Reconstruct every interior using $w^\star_K(E_K\Lambda + \lambda_{K,D})$. The interior equations hold by construction, the traces satisfy compatibility, and the reduced root enforces the remaining balance equations. Hence the reconstructed state is a full equilibrium. Uniqueness of each interior solution makes the correspondence one-to-one on the chosen branch.

If two local response maps agree on the admissible domain, substituting either map leaves every surrounding balance equation unchanged. Their reduced equilibrium sets and all exterior observables therefore agree. When a context selects one equilibrium, the selection convention must be included in the context. This proves the replacement statement. An interior observable also requires the corresponding interior decoder. The external response of a component does not by itself specify all of its internal observables.

This combines static condensation (Eftang & Patera, 2014) with local smooth elimination through the implicit-function theorem (Bonnans & Shapiro, 2000). Its role in the paper is to identify an exact reusable interface before choosing an approximation to that interface.

B.6 Covectors, metrics, and coordinate adapters

A boundary response t is dual to the trace λ, with virtual work $t^\top \delta\lambda$. For a change of trace coordinates $\lambda = P\mu$, the response becomes $P^\top t$. If R is a positive-definite trace metric, natural trace and dual coordinates are

$$z = R^{1/2}\lambda, \qquad r = R^{-1/2}t.$$

Their pairing remains $r^\top \delta z = t^\top \delta\lambda$. For an assembled Jacobian J, the normalized derivative is $R^{-1/2}JR^{-1/2}$. A direct neural covector model can have a nonsymmetric derivative. The exact variational HF operator can still have a symmetric tangent. These are different properties of the exact relation and its chosen evaluator.

B.7 Separating representation, evaluator, and reconstruction errors

Let $u^\star_{\text{fine}}$ be a finer reference, $u^\star_q$ the exact discrete solution in a selected port space, and $\widehat{\Lambda}_q$ the trace produced by an approximate evaluator. For a fixed output decoder $\mathcal{D}$, the triangle inequality gives

$$\begin{aligned}\left\|\mathcal{D}(\widehat{\Lambda}_q) - u^\star_{\text{fine}}\right\|_E \le \left\|\mathcal{D}(\widehat{\Lambda}_q) - \mathcal{D}(\Lambda^\star_q)\right\|_E + \left\|\mathcal{D}(\Lambda^\star_q) - u^\star_q\right\|_E \\ + \left\|u^\star_q - u^\star_{\text{fine}}\right\|_E . \end{aligned} \tag{24}$$

The first term measures movement of the decoder's input, the second its error at the selected-space equilibrium, and the third the difference between the selected discretization and finer reference. The experiment in Appendix A uses $u^\star_q$ as its reference, so the third term is absent from that comparison. Figure 4 studies an additional port and reference effect.

These terms are an upper-bound decomposition. Their vector errors can reinforce or cancel, so the observed squared error need not equal a sum of squared contributions. In particular, the field error at an approximate trace may fall slightly below the error at the HF trace. This explains why the HF-response reconstruction is used as a reference measurement, without interpreting it as a strict error floor.

B.8 Closure under further composition

Partition a composed group's trace into exterior variables λ_E and internal variables λ_I, with balance residuals g_E, g_I. If the internal balance $g_I(\lambda_E, \lambda_I) = 0$ determines a unique smooth function $\lambda_I = \phi(\lambda_E)$, the group exposes

$$\mathcal{T}_C(\lambda_E) = g_E(\lambda_E, \phi(\lambda_E)). \tag{25}$$

Differentiating internal balance gives $D\phi = -(D_I g_I)^{-1} D_E g_I$. Substitution yields

$$D\mathcal{T}_C = D_E g_E - D_I g_E (D_I g_I)^{-1} D_E g_I. \tag{26}$$

Thus composition returns another exterior relation, with the familiar Schur complement tangent. The statement holds on the regular branch where the eliminated block is invertible. It does not assume that a global nonlinear relation is exactly affine. Hierarchical solvers and port methods exploit this structure (Martinsson, 2015; Eftang & Patera, 2014).

B.9 Scope of the representation principle

The elliptic setting supplies a known response interface and explicit composition laws. Spatial equilibrium makes it possible to identify sufficiency on a fixed discrete trace space and a regular solution branch, then intervene on approximation without changing the requested output model. The contextual quotient describes the information question at this level. The experiments examine finite-port evaluators, particular trained models, and finite collections of compositions. Their distinct references and aggregation rules are given with each study.

Two extensions change the representation problem itself. In evolutionary systems, future interaction can depend on the component's state or history. A sufficient interface must then support state updates and preserve the future observables selected by admissible input histories. Predictive-state representations provide a precedent for defining state through observable consequences (Littman

et al., 2001). Multiple equilibrium branches similarly require a convention for state selection as part of the component or its context.

The second problem is interface discovery when no elimination argument supplies r. The quotient identifies the desired distinctions, but learning an executable representation of them requires informative contexts, suitable losses, and control of approximation error. The frozen-output diagnostic can test a candidate interface by measuring what changes when a fixed decoder is embedded in new compositions. The propagation study motivates evaluating those candidates through the geometry of their consumers. These are directions for extending the principle beyond the known-interface setting studied here.

C Propagation theory, actual errors, and controlled geometry

C.1 Local perturbation theorem and proof

The argument underlying Proposition 2 is a standard local root perturbation calculation (Bonnans & Shapiro, 2000; Deuflhard, 2004). We give a quantitative version for the constant response biases used in the experiment. All norms below are Euclidean norms in the normalized coordinates of (5).

Theorem 1 (A nearby root under a constant assembled bias). *Let $F(y^\star) = 0$, let $J_0 = DF(y^\star)$ be invertible, and suppose DF is L_F-Lipschitz on a closed ball containing $B_r(y^\star)$. Write $\beta = 1/\left\|J_0^{-1}\right\|_2 > 0$, $b = Bq$, and $r = 2\|b\|_2/\beta$. If $b \neq 0$ and $2L_F\|b\|_2/\beta^2 < 1$, then $F(y) + b = 0$ has a unique root in $B_r(y^\star)$. Its displacement satisfies*

$$\|\delta y\|_2 \le \frac{2\|b\|_2}{\beta}, \qquad \left\|\delta y + J_0^{-1}b\right\|_2 \le \frac{L_F}{2\beta}\|\delta y\|_2^2. \tag{27}$$

If $b = 0$, $y^\star$ remains a root and is locally unique.

Proof. Let $d = y - y^\star$ and define the Taylor remainder $N(d) = F(y^\star + d) - J_0 d$. On the ball, $\|N(d)\| \le L_F\|d\|^2/2$ and $\|DN(d)\| \le L_F\|d\|$. The perturbed equation is the fixed-point equation $d = \Psi(d) = -J_0^{-1}b - J_0^{-1}N(d)$. For $\|d\| \le r$,

$$\|\Psi(d)\| \le \frac{\|b\|}{\beta} + \frac{L_F r^2}{2\beta} = \frac{r}{2} + \frac{L_F r}{\beta}\frac{r}{2} < r.$$

The derivative norm of Ψ is at most $L_F r/\beta < 1$ throughout the ball. The contraction theorem gives a unique fixed point in that ball and the first bound. Rearranging the fixed-point equation and using the remainder estimate proves the second. For $b = 0$, local uniqueness follows from the inverse function theorem. □

Apply the theorem to $F = \bar{G}$, $J_0 = \bar{J}$, and $b = Bq$. For fixed geometry and $q \to 0$, the second bound is $O(\|q\|^2)$, which proves Proposition 2. Local uniqueness is on the stated ball and branch. The relative prediction error measures the accuracy of this approximation, independent of whether the Lipschitz constant has been explicitly estimated for a given case.

Corollary 1 (A reversed ranking). *Let v_1, v_2 be unit right singular directions of M with singular values $\sigma_1 > 2\sigma_2 > 0$. For sufficiently small positive ϵ, the perturbed root with $q_1 = \epsilon v_1$ has larger global displacement than the root with $q_2 = 2\epsilon v_2$, although $\|q_1\| = \|q_2\|/2$.*

Proof. The two displacement norms are $\epsilon\sigma_1 + O(\epsilon^2)$ and $2\epsilon\sigma_2 + O(\epsilon^2)$. Their difference has a strictly positive first-order coefficient. It is therefore positive for all sufficiently small $\epsilon > 0$. □

For a general trace-dependent learned error, both its value and derivative may change. The constant-bias experiment removes that derivative intervention. It tests how response-error direction alone changes the nearby nonlinear root.

Proposition 4 (Remainder for a state-dependent response defect). *Let $F(y^\star) = 0$, let $J_0 = DF(y^\star)$ be invertible, and let DF be L_F-Lipschitz on a convex neighborhood containing $y^\star$ and $\widehat{y}$. Suppose*

$F(\widehat{y}) + e(\widehat{y}) = 0$, e is continuously differentiable, and $\|De(y)\|_2 \le \eta$ on that neighborhood. For $\delta y = \widehat{y} - y^\star$ and $b = e(y^\star)$,

$$\left\|\delta y + J_0^{-1} b\right\|_2 \le \left\|J_0^{-1}\right\|_2 \left(\eta \|\delta y\|_2 + \frac{L_F}{2} \|\delta y\|_2^2\right). \tag{28}$$

For a fixed smooth defect shape $e = \epsilon e_0$, the local root supplied by the implicit-function theorem satisfies $\delta y = -\epsilon J_0^{-1} e_0(y^\star) + O(\epsilon^2)$ as $\epsilon \to 0$.

Proof. Taylor expansion gives $F(\widehat{y}) = J_0 \delta y + N(\delta y)$ with $\|N(\delta y)\| \le L_F \|\delta y\|^2 / 2$. The integral mean-value formula gives $\|e(\widehat{y}) - b\| \le \eta \|\delta y\|$. Substitute both expressions into the perturbed root equation and multiply by J_0^{-1}. The triangle inequality proves (28). For $e = \epsilon e_0$, invertibility of J_0 supplies a local smooth root with $\delta y = O(\epsilon)$. Then $\eta = O(\epsilon)$ and the bound is $O(\epsilon^2)$. □

This is a classical Taylor and implicit-function argument (Bonnans & Shapiro, 2000; Deuflhard, 2004) specialized to the response system. It separates variation of the learned defect from curvature of the reference equilibrium. The constant-bias intervention sets the first contribution to zero.

C.2 The consistent tangent of a learned response system

The derivative contribution in (28) can be included in the linear prediction. Let $F = \bar{G}$ denote the normalized HF residual and $e = \bar{d}$ the normalized assembled defect. The candidate system is $\widehat{F} = F + e$. At the HF root,

$$b = e(y^\star), \qquad \widehat{J} = D\widehat{F}(y^\star) = DF(y^\star) + De(y^\star).$$

This is the derivative of the complete evaluator and assembly, including the trace dependence of the local response approximation.

Proposition 5 (Prediction with the consistent evaluator tangent)**.** *Suppose $F(y^\star) = 0$ and $\widehat{F}(\widehat{y}) = 0$. Assume F and e are continuously differentiable on a convex neighborhood containing the segment between these states, $D\widehat{F}$ is L-Lipschitz there, and $\widehat{J}$ is invertible. Write $\kappa = \left\|\widehat{J}^{-1}\right\|_2$ and $\delta y = \widehat{y} - y^\star$. Then*

$$\left\|\delta y + \widehat{J}^{-1} b\right\|_2 \le \frac{\kappa L}{2} \|\delta y\|_2^2. \tag{29}$$

For $b \neq 0$, a sufficient condition for a unique candidate root in a ball about $y^\star$ is that the closed ball of radius $r = 2\kappa \|b\|_2$ is contained in this neighborhood and $2\kappa^2 L \|b\|_2 < 1$.

Proof. The integral Taylor formula gives

$$0 = b + \widehat{J}\delta y + \rho(\delta y), \qquad \rho(\delta y) = \int_0^1 [D\widehat{F}(y^\star + t\delta y) - \widehat{J}]\delta y \, dt.$$

Lipschitz continuity implies $\|\rho(\delta y)\|_2 \le L \|\delta y\|_2^2 / 2$. Multiplication by $\widehat{J}^{-1}$ proves the bound. For existence, the map $z \mapsto -\widehat{J}^{-1} b - \widehat{J}^{-1}\rho(z)$ has derivative norm at most $\kappa L r < 1$ on the stated ball. Its norm is at most $r/2 + \kappa L r^2/2 < r$. The contraction theorem gives the unique root. □

This is the classical Taylor and local-root argument (Bonnans & Shapiro, 2000; Deuflhard, 2004) applied to the candidate response system. It proves (7) on the stated neighborhood. For the constant intervention $e(y) = Bq$, $De = 0$ and the consistent tangent is the HF tangent. For a fixed smooth shape $e = \epsilon e_0$, both formulations have the same first-order limit as $\epsilon \to 0$. At a finite learned defect, including De accounts directly for its local variation.

A numerical reference y_0 has a finite residual $r_0 = F(y_0)$. If the candidate endpoint also has residual $r_1 = \widehat{F}(y_1)$, Taylor expansion uses $e(y_0) + r_0 - r_1$ in place of b. Omitting the endpoint residuals adds at most $\left\|D\widehat{F}(y_0)^{-1}(r_0 - r_1)\right\|_2$ to the bound. For an exact candidate root, $r_1 = 0$. The diagnostics record numerical reference residuals separately and retain solver-floor flags.

C.3 ACTUAL RESPONSE DEFECTS OF THE TRAINED MODELS

The follow-up diagnostic evaluates the three field-induced interfaces and the fixed learned relation on the 32 test cases of Appendix A. At each same-mesh HF equilibrium, evaluate

$$\delta t_K = \widehat{\mathcal{T}}_K(\lambda_K^\star) - \mathcal{T}_K^{\mathrm{HF}}(\lambda_K^\star), \qquad d = \sum_K E_K^\top \delta t_K.$$

With physical analytic anchor Hessians H_K and $R = \sum_K E_K^\top H_K E_K$, the three candidate predictors of displacement magnitude are

$$e_{\mathrm{local}} = \left(\frac{1}{16} \sum_K \delta t_K^\top H_K^{-1} \delta t_K \right)^{1/2}, \quad e_{\mathrm{assembly}} = \|d\|_{R^{-1}}, \quad e_{\mathrm{propagation}} = \left\| J_{\mathrm{HF}}^{-1} d \right\|_R. \tag{30}$$

The local norm uses the physical anchor without extra material-normalization factors. The measured endpoint is $\left\| \widehat{\Lambda} - \Lambda^\star \right\|_R$. It measures interface displacement. Field reconstruction has its additional error at the resulting trace.

Both candidate and reference solves retain zero initialization and the full nonsymmetrized Jacobian, 18 Newton iterations, tolerance 10^{-9}, and ten line-search trial steps. The predictor uses the HF Jacobian and is diagnostic. No fitted coefficient, adaptive corrector, or selected error direction enters it. The learned errors depend on the trace, as treated by Proposition 4.

For primary case summaries, each scalar metric first averages nodes17/33. Correlations are calculated separately for the field-induced and learned-relation evaluators. The combined field-induced comparison contains 96 seed–case pairs on 32 shared cases. The 192 unaveraged field-induced rows are reported separately. The fixed relation has 64 unique mesh–case observations, summarized into 32 cases. Its repeated evaluations in the decoder runs agree and are counted once.

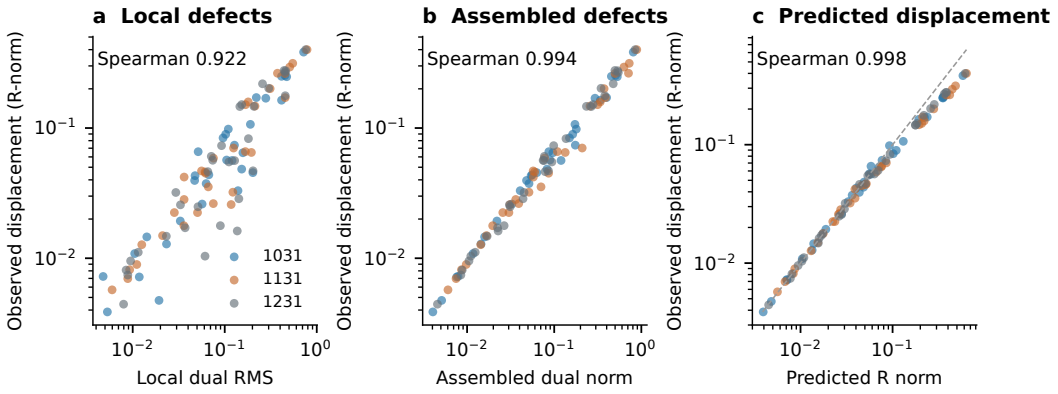

Figure 5: **Assembly and equilibrium sensitivity organize actual learned errors.** Each point is a field-decoder seed and test case after averaging the two meshes. Colors identify the three decoder seeds by their suffixes. The three predictors use the same observed nonlinear displacement. The final panel includes the identity line, with no fitted calibration.

The response defects combine before moving the global equilibrium. Most of the association gain in the pooled field-induced comparison comes from this assembly step. Propagation improves it further. Ordinary local error already correlates well within each decoder, while the combined comparison also varies the trained model. These are descriptive associations on the recorded cohort.

Relative prediction error compares the predicted and observed displacement vectors,

$$e_{\mathrm{pred}} = \frac{\left\| -\bar{J}_{\mathrm{HF}}^{-1} \bar{d} - \delta y \right\|_2}{\max(\|\delta y\|_2, 10^{-14})}.$$

The pooled case-level field-induced median is 8.13%, with maximum 58.85%. The predicted direction remains closely aligned, with minimum case-averaged cosine 0.996774. Thus the strong

Table 15: Actual-error association. Entries are Spearman / log-Pearson correlations. Per-seed and pooled seed–case rows average the meshes first. The raw-row diagnostic retains both meshes. Fixed-relation controls are deduplicated.

Model / unit	Count	Local	Assembled	Propagated
2026091031	32	0.8930 / 0.9388	0.9919 / 0.9958	0.9956 / 0.9971
2026091131	32	0.9487 / 0.9556	0.9916 / 0.9936	0.9993 / 0.9988
2026091231	32	0.9051 / 0.9007	0.9868 / 0.9974	0.9996 / 0.9988
All D seed–cases	96	0.9223 / 0.9306	0.9942 / 0.9953	0.9983 / 0.9980
All D raw rows	192	0.9194 / 0.9307	0.9937 / 0.9953	0.9985 / 0.9980
Fixed relation	32	0.9974 / 0.9985	0.9978 / 0.9995	1.0000 / 1.0000

ranking association coexists with finite-error magnitude deviations. The fixed learned relation has median 0.0238% and maximum 0.4614% case-level relative prediction error. Table 16 also gives all raw-row tails.

Table 16: Relative prediction error for actual learned response defects. The denominator is actual displacement. Mesh-averaged rows summarize case-level discrepancies; the unaveraged row retains their per-resolution maxima.

Model / unit	Median (%)	P95 (%)	Max (%)	Min cosine
2026091031	6.8816	41.8906	51.7189	0.997872
2026091131	10.8120	56.0384	58.8541	0.996774
2026091231	7.2380	38.2417	53.8213	0.998598
All D seed–cases	8.1299	51.5218	58.8541	0.996774
All D raw rows	8.0955	50.7183	60.8205	0.996428
Fixed relation	0.0238	0.2842	0.4614	1.000000

All diagnostic records are valid. A solver-floor flag compares predicted and actual displacements with ten times the HF residual correction, using a 10^{-14} floor. No row is flagged or excluded. Recorded assembly parity and linear-solve residuals check the construction; the packaged scalar records also satisfy the independent norm–cosine identity for prediction error. Full per-case data accompany the source.

C.4 Exchanging the Jacobian in the displacement prediction

The follow-up fixes the HF equilibrium, assembled field-induced defect and actual nonlinear displacement. It changes only the Jacobian used in $-\tilde{\bar{J}}^{-1}\bar{d}$. The four choices are HF, Analytic-2, Analytic-2+NN and the field-induced evaluator's own consistent tangent, all evaluated at the same HF trace. No tangent is symmetrized. The same 32 cubic cases and three frozen field decoders are used, with nodes17/33 paired within each case before per-seed statistics.

The consistent field-induced tangent gives median errors of 0.8605%, 0.9835% and 0.8334%. HF tangents give 6.8816%, 10.8120% and 7.2380%. The respective maximum case errors are 6.07%, 8.45% and 8.04% for the consistent tangent, compared with 51.72%, 58.85% and 53.82% for HF. Each Jacobian choice retains Spearman correlation above 0.995 with displacement magnitude. The main change is therefore the accuracy of the predicted vector.

All 768 diagnostic rows are finite, with no near-zero or solver-floor flags. Relative linear-system residuals are below 10^{-15}. The experiment supplies the HF state and response defect. It evaluates the finite-defect expansion in Proposition 5 and leaves the separate problem of estimating these quantities during deployment open. The scalar records and analysis code accompany the source package.

The dedicated cubic relation. We also evaluate the fixed relation with its own consistent tangent, using its own response defect and nonlinear displacement. The 64 frozen case–mesh records share relation seed 2026090617. The saved tangent and original displacement records refer to the same HF equilibrium, evaluator and normalized coordinates. Recomputing the HF prediction reproduces

Table 17: Predicting each field-induced evaluator's displacement with four Jacobians. Relative prediction-error quantiles (%) and Spearman correlation use 32 cases after mesh averaging. The denominator is actual displacement.

Decoder seed	Jacobian	Median (%)	P90 (%)	Max (%)	ρ_S
2026091031	HF	6.882	40.332	51.719	0.9956
2026091031	Analytic-2	7.389	40.395	51.821	0.9956
2026091031	Analytic-2+NN	7.013	40.349	51.746	0.9956
2026091031	Field-induced	0.861	5.346	6.067	0.9996
2026091131	HF	10.812	50.976	58.854	0.9993
2026091131	Analytic-2	10.818	51.124	58.993	0.9993
2026091131	Analytic-2+NN	10.813	51.025	58.884	0.9993
2026091131	Field-induced	0.983	6.095	8.453	0.9989
2026091231	HF	7.238	35.333	53.821	0.9996
2026091231	Analytic-2	7.249	35.562	53.913	0.9996
2026091231	Analytic-2+NN	7.238	35.407	53.844	0.9996
2026091231	Field-induced	0.833	6.241	8.037	0.9993

the previous relative prediction errors. Both predictors first average nodes17/33 within each of the 32 cases.

Table 18: Actual-error prediction for the dedicated cubic relation. Relative prediction-error quantiles are percentages of measured displacement. Own uses the relation's consistent Jacobian. Each row summarizes the same 32 mesh-paired cases.

Jacobian	Median (%)	P90 (%)	Max (%)	ρ_S
HF	0.023769	0.237367	0.461358	1.000
Own	0.000577	0.010497	0.030251	1.000

The consistent tangent reduces median relative prediction error from 0.023769% to 0.000577%, with maximum case error 0.030251%. Both predicted-magnitude rankings have Spearman correlation one. All 128 linear predictions are finite, with relative linear-system residuals below 9×10^{-16} and no solver-floor flags. The comparison uses the measured numerical endpoints and the response-defect convention of (7). Adding only the HF reference residual to this predictor gives median relative prediction error 0.002261% and maximum 3.315%. This sensitivity calculation targets an exact candidate root while the observed endpoint retains its own stopping residual. The endpoint-residual term above specifies the corresponding finite-solve correction. The source package contains the frozen matrices and their independent reconstruction.

C.5 Predicting actual response errors in sinh

The second-family diagnostic retains the full sinh reaction and the frozen models of Appendix A.6. For each of the 16 physical cases and two meshes, it computes one HF equilibrium, one shared learned-relation equilibrium and the three field-induced equilibria. This gives 160 unique nonlinear solves. Reconstruction through each fixed decoder reproduces all 288 earlier scalar measurements within 1.6×10^{-15} absolute error. All solves converge, all model parameters remain fixed and no response fallback is used.

At each HF equilibrium, we compute the local defects, their assembly and the metric in (30). Two fixed predictors use the same assembled defect and observed displacement. One uses the HF Jacobian and the other uses the candidate evaluator's consistent Jacobian, both evaluated at the HF trace without symmetrization. The 128 unique evaluator–case–mesh records give 256 prediction rows. Each scalar metric first averages nodes17/33 within a test case. Quantiles and correlations then use the 16 cases, separately for each field decoder and for the single shared relation.

The consistent tangent gives median relative prediction errors of 0.00509%, 0.00941% and 0.00720% for the three field-induced evaluators, and 0.11482% for the learned relation. Their maximum case errors are 0.08114%, 0.06937%, 0.05189% and 1.05465%, respectively. HF tangents give larger median relative prediction errors for every evaluator (Table 19). All eight predicted-

Table 19: Actual-error prediction in sinh. Relative prediction-error quantiles are percentages of the measured nonlinear displacement norm. Own denotes the candidate's consistent Jacobian. Correlations compare predicted and actual displacement magnitudes after paired mesh averaging over 16 cases.

Evaluator	Seed	Jacobian	Median (%)	P90 (%)	Max (%)	ρ_S
Field-induced	2026091251	HF	0.3605	1.2880	3.4754	1.000
Field-induced	2026091251	Own	0.0051	0.0484	0.0811	1.000
Field-induced	2026091252	HF	0.4371	1.1937	1.4647	1.000
Field-induced	2026091252	Own	0.0094	0.0468	0.0694	1.000
Field-induced	2026091253	HF	0.4671	1.1718	2.0467	1.000
Field-induced	2026091253	Own	0.0072	0.0383	0.0519	1.000
Learned relation	2026090837	HF	3.5797	5.2902	5.9573	1.000
Learned relation	2026090837	Own	0.1148	0.6021	1.0546	1.000

magnitude rankings have Spearman correlation one. Local dual-error correlations are 0.903–0.950 within the field decoders and 0.991 for the learned relation. Assembly raises these to 0.994–1.000 and 1.000. Local error is already informative here, and explicit composition improves the prediction of both magnitude and direction.

Every record is finite and no displacement is flagged near the solver floor. Saved vectors, metrics, Jacobians and local defects permit independent reconstruction of the predictions. The experiment supplies the HF root and response defect, with all models frozen. It verifies actual-error propagation in the second family. The independent direction intervention in Appendix C uses cubic physics.

C.6 Fixed physical cohort and metrics

The one-dimensional cubic component response has two endpoint covector entries. The physical cohort contains 72 historical assemblies, once each, formed from three component counts $N \in \{8, 16, 32\}$, four patterns (alternating, localized, repeated-block, and high-contrast), and six boundary replicates. The catalog reconstructs the recorded physical inputs using seed 64001. Historical training seeds share those same assemblies and are not new units in this experiment. No checkpoint is loaded.

Let A assemble the $2N$ local outward responses into the $N-1$ free interior balances. The local norm is exactly (8). The diagonal matrix L has two entries $\sqrt{Na_K/h_K}$ per component, so $\delta t = Lq$. The global metric R is the assembled linear-diffusion energy matrix with local blocks

$$\frac{a_K}{h_K}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

restricted to free traces. Dirichlet endpoints make R positive definite. We form $M = \bar{J}^{-1}R^{-1/2}AL$ using the same physical norms as the measured local and global errors. Singular vectors of an unweighted raw matrix would specify a different intervention.

C.7 Directions, magnitudes, and nonlinear solves

The sensitive direction is the right singular vector for the largest singular value. The comparison direction is the median nonzero right singular vector, using a rank tolerance of 10^{-12} times the largest singular value. For descending singular values and rank r_M, the zero-based median index is $\lfloor (r_M - 1)/2 \rfloor$. Each sign is fixed by making its largest absolute coordinate positive. These rules apply to every case without selecting directions from the perturbed outcomes.

For each $\epsilon \in \{10^{-5}, 3\times10^{-5}, 10^{-4}, 3\times10^{-4}, 10^{-3}\}$, the three conditions are $\epsilon v_{\max}$, $\epsilon v_{\mathrm{mid}}$, and $2\epsilon v_{\mathrm{mid}}$. The implementation adds Lq inside each actual local HF response call. This bias is constant in the trace. The exact nonlinear HF tangent returned at that trace is therefore unchanged by the intervention. The local solver uses 32 substeps per component.

The reference equilibrium starts from linear interpolation between the external Dirichlet endpoints. Every perturbed solve starts from that same case's HF equilibrium, then recomputes nonlinear responses and tangents. Newton allows 20 iterations and a full step followed by ten halving levels. Its

stopping rule is a normalized residual norm of 10^{-11}. No linear prediction substitutes for a nonlinear solve. All 72 references, all 72 zero-bias controls, and all 1,080 perturbed solves converge. The run takes 121.531 seconds on one CPU thread, without GPU allocation.

C.8 Measurements and aggregation

For each row we record the injected local norm, the nonlinear displacement $\|\delta\Lambda\|_R$, predicted magnitude $\|Mq\|_2$, relative prediction error

$$e_{\text{pred}} = \frac{\|\delta y + Mq\|_2}{\max(\|\delta y\|_2, 10^{-14})},$$

and cosine similarity between the measured and predicted displacement vectors. The denominator is the observed displacement norm, as in the trained-model diagnostics. The global-error ratio at equal local error divides the sensitive displacement by the median-direction displacement at the same ϵ. The ranking-reversal ratio instead divides by the doubled-median displacement. Magnitudes and directions are paired measurements, and test cases receive equal weight.

At the primary magnitude 10^{-4}, the all-case median equal-norm ratio is 7.835037. The median ranking-reversal ratio is 3.917518. Every spectrum predicts the reversal, and every nonlinear comparison exhibits it. For the small-error summary, each case first takes its worst relative prediction error over all three conditions and the three base magnitudes at most 10^{-4}. The median of those 72 maxima is 0.004931%, the 95th percentile is 0.036431%, and the maximum is 0.053276%.

Table 20: Depth-stratified nonlinear results at $\epsilon = 10^{-4}$. The equal-norm and half-norm columns compare within the same assembly.

N	Cases	Median ratio	Max ratio	Half norm ratio	Reversals
8	24	3.925234	4.448228	1.962617	24/24
16	24	7.835037	8.869602	3.917518	24/24
32	24	15.671001	17.730190	7.835501	24/24

Table 21: Complete magnitude sweep. Ratios are medians across 72 cases. Relative prediction error is the maximum over all 216 rows at that base magnitude.

ϵ	Equal norm ratio	Half norm ratio	Max prediction error (%)	Converged
1e-05	7.8349	3.9175	0.005342	216/216
3e-05	7.8350	3.9175	0.016017	216/216
0.0001	7.8350	3.9175	0.053276	216/216
0.0003	7.8352	3.9176	0.158860	216/216
0.001	7.8357	3.9179	0.518061	216/216

Recorded checks verify local-norm agreement, exact assembly of the injected bias, and tangent identity at the reference and perturbed endpoint. The largest relative local-norm mismatch is 9.04×10^{-16}. The largest final normalized residual is 9.92×10^{-12}. Independent reaggregation of the saved matrices reproduces all 1,080 prediction magnitudes and local norms. The controlled singular directions test the effect of error direction and its quantitative prediction. The actual-error diagnostic above supplies the complementary observations from trained interfaces.

Table 22: All 72 controlled assemblies. Direction ratios use base magnitude 10^{-4}. Relative prediction error is the worst across all arms and base magnitudes at most 10^{-4}.

Case	N	Pattern	Equal norm	Half norm	Prediction error (%)
01	8	alternating	3.9242	1.9621	0.00319
02	8	alternating	3.9367	1.9684	0.00282
03	8	alternating	3.8693	1.9347	0.00500
04	8	alternating	3.9262	1.9631	0.00293

Case	N	Pattern	Equal norm	Half norm	Prediction error (%)
05	8	alternating	3.8914	1.9457	0.00428
06	8	alternating	3.9654	1.9827	0.00076
07	8	localized	4.4478	2.2239	0.00045
08	8	localized	4.4116	2.2058	0.00315
09	8	localized	4.3274	2.1637	0.00548
10	8	localized	4.4422	2.2211	0.00145
11	8	localized	4.4419	2.2210	0.00164
12	8	localized	4.4482	2.2241	0.00033
13	8	repeated block	3.8449	1.9224	0.00262
14	8	repeated block	3.8413	1.9207	0.00288
15	8	repeated block	3.8792	1.9396	0.00036
16	8	repeated block	3.8755	1.9377	0.00066
17	8	repeated block	3.8788	1.9394	0.00049
18	8	repeated block	3.8407	1.9203	0.00256
19	8	high contrast	4.0239	2.0119	0.00087
20	8	high contrast	3.9682	1.9841	0.00368
21	8	high contrast	4.0093	2.0046	0.00190
22	8	high contrast	3.8759	1.9379	0.00567
23	8	high contrast	3.9209	1.9605	0.00483
24	8	high contrast	3.9155	1.9577	0.00492
25	16	alternating	7.8706	3.9353	0.00107
26	16	alternating	7.8270	3.9135	0.00470
27	16	alternating	7.6462	3.8231	0.01463
28	16	alternating	7.8555	3.9278	0.00432
29	16	alternating	7.8068	3.9034	0.00841
30	16	alternating	7.7151	3.8576	0.01250
31	16	localized	8.8165	4.4083	0.00756
32	16	localized	8.8402	4.4201	0.00593
33	16	localized	8.8131	4.4065	0.00780
34	16	localized	8.7524	4.3762	0.01111
35	16	localized	8.8507	4.4254	0.00390
36	16	localized	8.8696	4.4348	0.00237
37	16	repeated block	7.6233	3.8117	0.00431
38	16	repeated block	7.6027	3.8014	0.00576
39	16	repeated block	7.5828	3.7914	0.00681
40	16	repeated block	7.4834	3.7417	0.01096
41	16	repeated block	7.5747	3.7873	0.00667
42	16	repeated block	7.4650	3.7325	0.01168
43	16	high contrast	7.7609	3.8804	0.01177
44	16	high contrast	7.8431	3.9216	0.00343
45	16	high contrast	7.8485	3.9242	0.00388
46	16	high contrast	7.8899	3.9449	0.00345
47	16	high contrast	7.7768	3.8884	0.01108
48	16	high contrast	7.9005	3.9503	0.00073
49	32	alternating	15.7103	7.8552	0.00349
50	32	alternating	15.4223	7.7112	0.03361
51	32	alternating	14.9126	7.4563	0.05321
52	32	alternating	14.9098	7.4549	0.05328
53	32	alternating	15.2094	7.6047	0.04346
54	32	alternating	15.7151	7.8576	0.00410
55	32	localized	17.5817	8.7908	0.02504
56	32	localized	17.6953	8.8477	0.01024
57	32	localized	17.7302	8.8651	0.00482
58	32	localized	17.6500	8.8250	0.01888
59	32	localized	17.7269	8.8635	0.00661
60	32	localized	17.3256	8.6628	0.03988
61	32	repeated block	15.2308	7.6154	0.00298
62	32	repeated block	15.1820	7.5910	0.01048
63	32	repeated block	15.2243	7.6121	0.00721
64	32	repeated block	15.2356	7.6178	0.00386
65	32	repeated block	15.1773	7.5887	0.01172
66	32	repeated block	15.0061	7.5030	0.02586
67	32	high contrast	15.6731	7.8366	0.00908

Case	N	Pattern	Equal norm	Half norm	Prediction error (%)
68	32	high contrast	15.5252	7.7626	0.02716
69	32	high contrast	15.7206	7.8603	0.00495
70	32	high contrast	15.7019	7.8509	0.00937
71	32	high contrast	15.6898	7.8449	0.01061
72	32	high contrast	15.6689	7.8344	0.00703

D Analytic and neural evaluation of the local relation

D.1 The cubic teacher and linear anchor

The scalar teacher uses Q1 finite elements and tensor-product quadrature on the reference square. For fixed parameters and a prescribed port trace, it solves the interior nonlinear equations in double precision. The interior residual RMS tolerance is 10^{-12} with at most 30 Newton updates. Boundary responses are weak residual covectors pulled back by the port basis. The HF tangent is the exact nonlinear interior Schur complement. Its variational symmetry is enforced in the HF backend.

The linear operator is $A_0 = aK + h^2 cM$. Its forced solution with prescribed trace is v. Let G_0 solve the dimensionless, homogeneous-trace linear interior problem and set $\varepsilon = \gamma h^2/a$. For the quartic perturbation, $w = G_0(v^3)$ is the first zero-trace correction in $u = v - \varepsilon w + O(\varepsilon^2)$. The normalized reduced actions retained by the analytic evaluators are

$$S^{[2]} = S_0 + \frac{\varepsilon}{4}\int v^4 - \frac{\varepsilon^2}{2}\int v^3 w, \tag{31}$$

$$S^{[3]} = S^{[2]} + \frac{3\varepsilon^3}{2}\int v^2 w^2. \tag{32}$$

All integrals use the same discrete quadrature as the teacher. Analytic-k computes $D_\lambda(aS^{[k]})$ and its trace derivative. It uses cached linear liftings and linear interior solves. The order labels refer to this expansion for cubic, not to Q1 polynomial degree.

To see the coefficients, write the normalized energy as $S_0(u) + \varepsilon\mathcal{N}(u)$ with $\mathcal{N}(u) = \int u^4/4$. The first interior correction solves $A_{0,I}(-w) + \mathcal{N}_I'(v) = 0$. Substituting the stationary expansion into the energy cancels the unknown second interior correction. The second-order coefficient is $-\langle \mathcal{N}_I'(v), G_0\mathcal{N}_I'(v)\rangle/2$. The third-order coefficient is $\mathcal{N}''(v)[w,w]/2 = 3\int v^2 w^2/2$. These are standard constrained perturbation coefficients (Bonnans & Shapiro, 2000) specialized to the cubic local problem.

D.2 Natural coordinates and the normalized remainder

Let H_{anc} be the condensed boundary matrix of the linear anchor and $R_K = H_{\mathrm{anc}}/a$. Symmetric spectral square roots define $z = R_K^{1/2}\lambda$. For the HF-minus-Analytic-2 response and tangent, set

$$r = R_K^{-1/2}(t_{\mathrm{HF}} - t_{[2]}), \tag{33}$$

$$J_r = R_K^{-1/2}(H_{\mathrm{HF}} - H_{[2]})R_K^{-1/2}. \tag{34}$$

At quadrature points the linear anchor has values $u_f + B_z z$. Its amplitude and the perturbative response scale are

$$A_K(z) = \left(\frac{\sum_i w_i (u_{f,i} + (B_z z)_i)^4}{\sum_i w_i}\right)^{1/4}, \qquad s_K = \frac{h^6\gamma^3}{a^2}A_K^7. \tag{35}$$

The implementation floors the fourth-power average at 10^{-32} and uses $s_{\mathrm{safe}} = \max(s_K, 10^{-30})$ for labels. The normalized targets are

$$q^\star = \frac{r}{s_{\mathrm{safe}}}, \qquad J_q^\star = \frac{J_r}{s_{\mathrm{safe}}} - \frac{r(\nabla_z s_K)^\top}{s_{\mathrm{safe}}^2}. \tag{36}$$

The derivative includes the scale product rule. In this appendix $q^\star$ is a training target. The controlled bias coordinate q in Appendix C is a separate mathematical object.

The network returns a direct natural-coordinate vector. With frozen output scale σ_Q, its physical covector is $Q_\theta^{\mathrm{phys}} = R_K^{1/2}\sigma_Q Q_\theta(z,p)$. The activation is

$$\omega(s) = \mathrm{sigmoid}\left(\frac{\log(s+10^{-30}) - \log(s_0+10^{-30})}{w_0}\right).$$

For the cubic seed used in the frozen-output intervention, $s_0 = 3.645504393922538 \times 10^{-8}$ and $w_0 = 1$. Calibration uses fresh isolated local probes after training, with weights unchanged. The deployed tangent differentiates the entire $t_{[2]} + \omega(s)sQ_\theta^{\mathrm{phys}}$ expression, including derivatives of the scale and activation. It is not symmetrized.

D.3 Cubic neural architecture, data, and optimization

The network combines a linear projection of 16 normalized trace coefficients with a two-layer, width-64 context encoder. Three width-64 SiLU layers and a linear 16-dimensional head produce the remainder, for 19,216 parameters. The physical context is

$$(h^2c/a,\ h^2\gamma/a,\ h,\ a,\ h^2f_0/a,\ h^2f_x/a,\ h^2f_y/a).$$

The first four context entries use training means and standard deviations. The three load entries share a training RMS. A pooled training RMS normalizes the trace. The physical parameter box is

$$a \in [0.7, 1.8], \quad c \in [1, 8], \quad \gamma \in [0.25, 90], \quad f_0 \in [-40, 40], \quad f_x, f_y \in [-15, 15], \quad h = 0.25.$$

Component width is fixed throughout these main training runs. Each resolution contributes 192 training, 48 validation, and 64 local test parameter configurations, with four probes per configuration. Configurations are split before their correlated probes are generated. Gaussian directions are normalized to unit length in natural trace coordinates, with radii sampled log-uniformly on $[0.05, 1.2]$. The same model alternates the two training resolutions.

The value loss fits normalized remainder labels, while a physical JVP loss follows the differentiated response. For natural unit directions v,

$$\widehat{J}_r v = s_K D_z(\sigma_Q Q_\theta)v + \sigma_Q Q_\theta(\nabla_z s_K)^\top v.$$

The value term is mean squared error divided by σ_Q^2. The JVP term averages the following normalized error over four directions and 32 examples per update,

$$\frac{\left\|\widehat{J}_r v - J_r v\right\|_2^2}{\|J_r v\|_2^2 + \tau_J^2}, \qquad \tau_J = \max\left(\frac{0.1\,\mathrm{median}\,\|J_r\|_F}{\sqrt{16}}, 10^{-14}\right).$$

The median defining the denominator floor uses eligible training responses. Training uses $\mathcal{L}_V + \lambda_J\mathcal{L}_J$ with $\lambda_J \in \{0.25, 1\}$, 5,000 updates, batch 128, AdamW learning rate 10^{-3}, weight decay 10^{-4}, and gradient clipping at two. Derivative supervision follows the general principle of Sobolev training (Czarnecki et al., 2017).

Validation at steps 1, 250, and subsequent multiples of 250 selects the checkpoint with the lowest maximum-across-resolution median physical-tail Jacobian error. The three training seeds are 2026090617, 2026090717, and 2026090817. The first is used alone in the new frozen-output intervention. The earlier breadth summary retains all three frozen seed models. No global equilibrium or layout appears in these local training losses.

D.4 An implementation view

Figures 6 and 7 summarize the recorded training and the implementation. Analytic subtraction, normalization, numerical assembly, and parent execution are choices for computing and using the reusable interface.

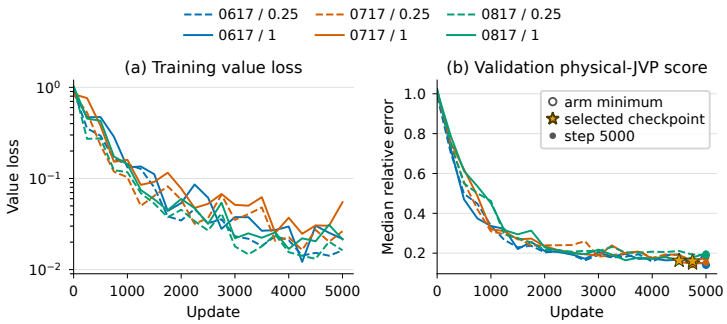

Figure 6: Recorded cubic training histories for six seed and JVP-weight runs. Open circles mark selected checkpoints, stars mark the three frozen primary models, and dots mark final iterates.

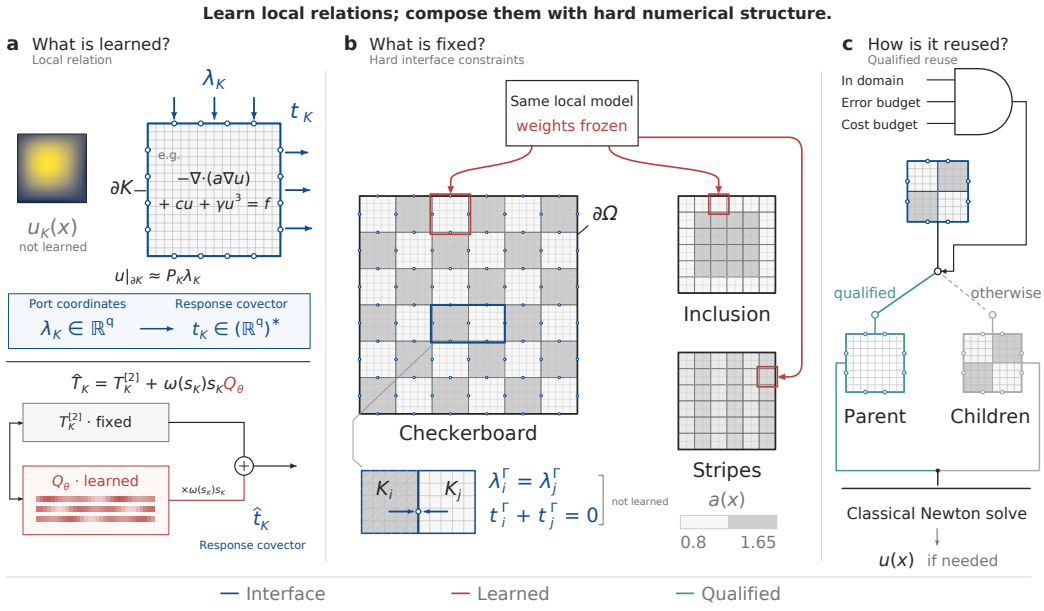

Figure 7: An analytic-plus-neural realization. Isolated training supplies a local response remainder, frozen weights enter unseen assemblies, and a composed group can expose a parent interface. Appendix F specifies the one-level parent execution and reference-based verification.

E SUPPORTING TARGET, PHYSICS, AND INTERFACE STUDIES

E.1 THE EARLIER TARGET COMPARISON

The target screen uses 32 existing cubic test cases at nodes17/33. The local labels contain 192 training, 48 validation, and 64 test parameter configurations per resolution with four probes each. Four newly trained controls use seed 2026090811, 5,000 updates, batch 32, learning rate 10^{-3}, weight decay 10^{-4}, gradient clipping at two, and validation every 250 updates. The vector controls have 23 inputs, two width-64 SiLU hidden layers, and 16 outputs. They learn a direct response, a raw natural-coordinate remainder, or a normalized remainder value. The normalized-value arm retains the exact scale product rule in its evaluation tangent. The main Analytic-2+NN model is inherited from the separate value-and-JVP training in Appendix D.

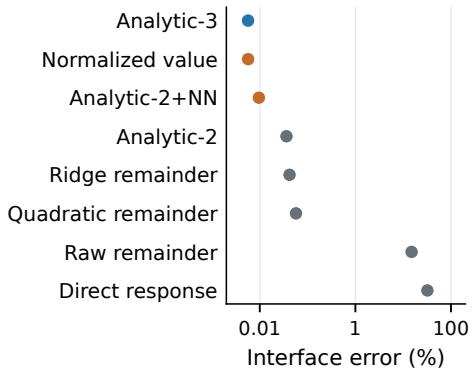

Figure 8: **The parameterization of a relation matters.** Mean interface errors over 32 cubic cases at 33 nodes per component axis with one composer and same-port HF. The complete table retains the historical field control.

The earlier field control is a width-64 coordinate network with two hidden layers, 25 inputs and a scalar interior output, with exact prescribed boundary values. Its loss combines normalized interior-field and weak-flux residual terms. This is a different model from the strengthened width-256 reconstructor $\mathcal{D}$. Three of its 33-node solves fail to converge. Their recorded errors remain included in the complete table below. Figure 8 displays only response-interface evaluators. The field-versus-interface experiment uses the fixed stronger decoder in Figure 1. Ridge and degree-two polynomial controls use the same isolated labels with a validation-selected regularization grid $\{10^{-8}, 10^{-6}, 10^{-4}, 10^{-2}, 1\}$.

Table 23: Complete earlier 33-node target comparison. Interface errors use same-port HF. All 32 cases remain in each mean.

Target / evaluator	Mean interface error (%)	Nonconverged rows
Analytic-2	0.036147	0
Analytic-3	0.005686	0
HF	0.000000	0
Analytic-2+NN	0.009589	0
Direct response	32.084557	0
Raw remainder	15.005123	0
Normalized value	0.005710	0
Local field	62.976773	3
Ridge remainder	0.041970	0
Quadratic remainder	0.057193	0

The direct raw response mean is 32.084557%, the raw remainder mean is 15.005123%, and the inherited Analytic-2+NN mean is 0.009589%. Their ratio exceeds three orders of magnitude. The normalized-value control reaches 0.005710%. These complete parameterizations have different models and objectives, so this screen supports a design choice under recorded budgets. The fixed-reconstructor intervention in Appendix A isolates a different comparison.

E.2 THREE-FAMILY TRAINING AND RESULTS

The cubic family uses the realization in Appendix D. Its earlier frozen-assembly summary averages 32 test cases at two mesh resolutions, with three independently trained models. Table 24 retains the separate seed results. The earlier cubic field metric uses same-mesh full-nodal HF, while its interface metric uses same-port HF. This differs from the same-port field reference of the new intervention.

Nonlinear elasticity. In Mandel strain coordinates, the constitutive response is

$$\sigma = C\varepsilon + \alpha \|\varepsilon\|_2^2 \varepsilon, \qquad C = 2\mu I + \lambda_L [1,1,0]^\top [1,1,0].$$

The backend uses P1 triangles with sparse Newton tangents. The parameter box is $\mu, \lambda_L \in [0.8, 2]$, $\alpha \in [50, 200]$, and body-force components in $[-0.15, 0.15]$, with $h = 0.25$. A 32-dimensional vector port is projected away from three rigid motions. The network has 37 inputs, two width-128 Tanh layers, and 32 outputs. The projected trace is normalized by $R_0 = 0.004$, context by $(2, 2, 200, 0.15, 0.15)$, and remainder by $\alpha \|P\lambda\|_2^3 / h^2$. Each seed has 1,024 isolated probes across nodes17/33, with trace radii between 5% and 100% of R_0. Adam uses learning rate 10^{-3}, batch 64, and 1,500 updates. The fixed final iterate is used, with the recorded 80/20 split retained for training diagnostics. Seeds are 2026090817, 2026090818, and 2026090819.

The 16 global cases contain four patterns and four repeats. Newton allows 25 iterations and 12 line-search halvings. The learned evaluator uses local HF when the projected trace exceeds R_0. All solves converge. There are 14 fallback calls per seed, all on one stripe case, with seven of 128 local calls at each mesh. These two row fractions are 5.46875%, exceeding the original 5% budget. The other 15 cases require no fallback. All 32 case-resolution rows per seed meet the one-percent interface and field accuracy criteria, and 30 of 32 also meet the original fallback budget. All rows contribute to the reported errors. The field metric integrates the linear Mandel-strain energy against the family HF reference at the same port and mesh.

Sinh reaction. The scalar equation is

$$-\nabla \cdot (a \nabla u) + cu + \gamma(\sinh u - u) = f.$$

The Q1 backend uses nine-point quadrature, $a \in [0.8, 2]$, $c \in [0.5, 2]$, $\gamma \in [2, 8]$, $f \in [12, 36]$, and $h = 0.25$. The 16-dimensional trace is scaled by $R_0 = 5$ and context by $(2, 2, 8, 36)$. The remainder scale is $\gamma h^2 \|\lambda\|_2^3 / 6$. The network has 20 inputs, two width-128 Tanh layers, and 16 outputs. Each seed uses 1,024 probes crossing resolution and trace-type strata, split 819/205 for training and validation. Adam uses learning rate 10^{-3}, batch 64, and 1,500 updates. Validation every 100 steps selects the minimum MSE with earliest tie. Seeds are 2026090837, 2026090838, and 2026090839. The 16 global cases are paired across the two meshes. No local fallback occurs. The field norm integrates $a \|\nabla v\|^2 + cv^2$ against the same-port family HF reference. The learned means improve Analytic-2 while remaining above one percent for this family, and Analytic-3 is more accurate.

Table 24: Recorded primary learned seed results (%). Cubic rows retain individual resolutions. Elasticity and sinh rows average their paired resolutions within cases. The field reference is family-specific as defined in the text.

Family	Seed	Nodes	Interface (%)	Field (%)
Cubic	2026090617	17	0.009339	0.377077
Cubic	2026090617	33	0.009589	0.442285
Cubic	2026090717	17	0.009273	0.377011
Cubic	2026090717	33	0.009497	0.442220
Cubic	2026090817	17	0.007130	0.376687
Cubic	2026090817	33	0.007355	0.441922
Elasticity	2026090817	17/33	0.167500	0.162164
Elasticity	2026090818	17/33	0.174513	0.168946
Elasticity	2026090819	17/33	0.184456	0.178571
Sinh	2026090837	17/33	1.449678	1.400610
Sinh	2026090838	17/33	1.402234	1.354780
Sinh	2026090839	17/33	1.511141	1.460091

Family-specific analytic orders. For elasticity and sinh, the analytic formulas expand in boundary trace about the loaded zero-trace HF solution u_0. Let $F(u)$ be the discrete residual, $A_0 = DF(u_0)$, and L_0 the equilibrium lifting satisfying $(A_0 L_0)_I = 0$. Set $v = L_0 \lambda$ and solve the zero-trace correction $A_{0,II} w_I = -[D^2 F(u_0)[v, v]]_I / 2$. With anchor response t_0 and $S = L_0^\top A_0 L_0$, the

responses are

$$\mathcal{T}^{[2]}(\lambda) = t_0 + S\lambda + \tfrac{1}{2}L_0^\top D^2F(u_0)[v,v], \tag{37}$$

$$\mathcal{T}^{[3]}(\lambda) = \mathcal{T}^{[2]}(\lambda) + L_0^\top \left(D^2F(u_0)[v,w] + \tfrac{1}{6}D^3F(u_0)[v,v,v]\right). \tag{38}$$

The identity $L_0^\top A_0 w = 0$ removes the explicit second-order interior correction from the second-order response. Differentiation supplies each tangent. Full sinh/cosh and constitutive derivatives are retained. Thus the order labels have an explicit family-specific meaning, while all evaluators share a boundary-response interface.

E.3 Port transfer, field references, and classical calibration

The 65-node transfer retains the same trained cubic model, fixed component width $h = 0.25$, and 16-dimensional port. There are 16 new confirmation cases. Relative interface-trace errors are measured against same-port HF. Relative energy-norm field errors use a finer Q1 finite-element reference, FEM129. The label FEMn means n nodal points per local component axis, so FEM33, FEM65, and FEM129 are mesh-resolution names. The assembled 4×4 domain shares nodes at component boundaries.

On the 16 transfer cases, Analytic-2+NN has maximum relative interface error 0.0981% and mean field error about 0.726%. Twelve cases are within the one-percent field criterion. FEM65 gives about 0.502% mean field error and meets the criterion on all 16. These field comparisons use the same FEM129 reference. A separate eight-case port audit raises the port dimension from 16 to 24 and 32 with HF and Analytic-3. Both reach the one-percent field criterion on all eight cases at the richer ports. The field mean at $q = 32$ is about 0.53%. This audit changes interface capacity using numerical evaluators.

An earlier classical calibration uses 32 cases and FEM129 as reference to select FEM65. FEM65 reaches 0.5008% mean field error and satisfies the one-percent condition on all 32 cases. It takes about 1.69 seconds on average in that calibration. The separate transfer cohort records about 1.78 seconds for FEM setup, solve, and error evaluation. These are different cohorts and timing scopes. The target-screen field reference is FEM65, so its field errors are excluded from the FEM129 calibration and transfer ranking. Meaningful speed comparisons require matched accuracy criteria and clear timing scopes (McGreivy & Hakim, 2024).

E.4 Local trace-amplitude coverage

A recorded diagnostic replays the earlier 32 cubic test cases at nodes17/33 with the frozen relation evaluator. All 64 solves converge. Every converged local trace radius lies inside the isolated-training support $[0.05, 1.2]$ in natural coordinates. The pooled converged minimum, median, and maximum are approximately 0.138, 0.448, and 1.012. Nonzero solver queries exceed the support on 0.130% of queries when fractions are averaged equally over cases. Directly pooling queries gives 0.167%. All excursions arise in the first line-search trial of one inclusion case, with largest radius about 1.389. Zero initialization is recorded separately. Training support is reconstructed from the frozen sampler and seeds.

E.5 A boundary case in nonlinear diffusion

The supplementary p-Laplacian family solves

$$-\nabla\cdot\left(a(\delta^2 + \|\nabla u\|_2^2)^{(p-2)/2}\nabla u\right) + cu = f, \qquad p = 3.$$

The parameter ranges are $\delta \in [0.8, 1.2]$, $a \in [0.8, 2]$, $c \in [0.5, 2]$, and $f \in [1, 4]$, with $h = 0.25$, a 16-dimensional port, and trace radius 0.8. The corrected value-only training uses the same 20–128–128–16 Tanh network, 1,024 probes with an 819/205 split, Adam learning rate 10^{-3}, batch 64, and 1,500 updates. Seeds are 2026090907, 2026090908, and 2026090909. Sixteen test cases are paired across nodes17/33.

Analytic-2 and Analytic-3 have mean interface errors 7.011237% and 0.391718%. The three value-trained learned means are 7.282168%, 6.823291%, and 9.455242%. A separate original-scale physical-JVP protocol gives 6.930356%, 6.983811%, and 6.718570%. No learned seed meets the

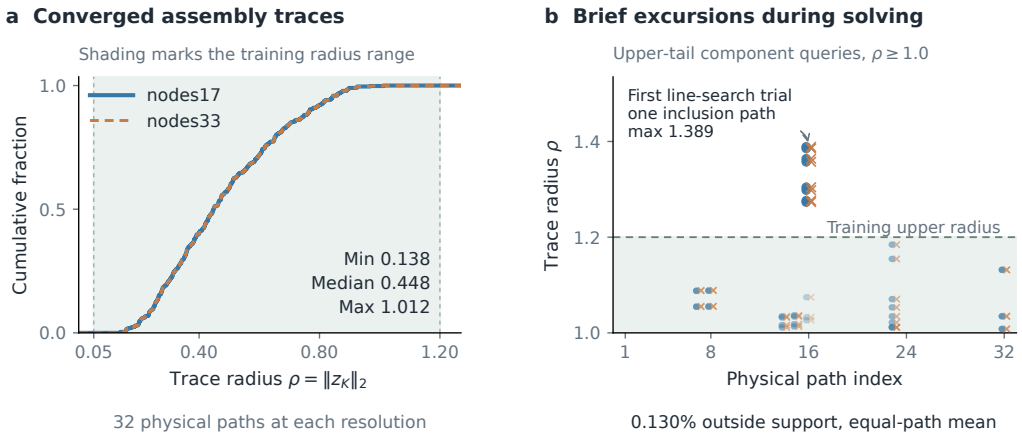

Figure 9: Amplitude coverage in the earlier cubic assembly cohort. Converged traces remain in the isolated-training radius support, while the few solver-time excursions are transient. This diagnostic concerns natural-coordinate amplitude coverage. It does not certify interpolation in the full trace-direction space. The cohort is separate from the new frozen-output confirmation.

one-percent criterion on any of the 32 case-resolution rows, while Analytic-3 meets the joint criterion on 24. These results do not establish a stable learned gain over Analytic-2 in this family. They retain a useful distinction between a sufficient numerical interface and the quality of a particular learned evaluator for that interface.

F Computational consequences and further reuse

F.1 Replacing repeated local solves

The earlier batched benchmark exchanges local response evaluators inside the same global interface solver. It uses the fixed cubic seed 2026090617 model, the same activation as Appendix D, a 4×4 assembly and $q = 16$. Newton allows 14 steps, residual tolerance 10^{-9}, and eight line-search steps. Batch sizes are $1, 8, 32$ for nodes17 and $1, 4, 8$ for nodes33. Each batch takes the first B problems from a fixed ordered 23-case catalog, cycling when $B > 23$. Evaluators see identical inputs at each batch size. The case mixture can change between batch sizes.

Each reported time is the median of three synchronized runs after one warmup, divided by the number of PDEs in the batch. The local-cell microbatch sizes are 32 for nodes17 and eight for nodes33. Timing includes interface residual assembly, Newton updates, and line search. It excludes trace-independent setup, field reconstruction, label generation, and training. Computation is FP64 with explicit tangents on the shared RTX 5090 server. The run receipt identifies its GPU lane, and contemporaneous host records supply the device model.

Table 25: All six measured batch configurations. Method columns are warmed interface-solve seconds per PDE. HF/NN divides HF time by Analytic-2+NN time.

Nodes	Batch	Analytic-2	Analytic-3	Analytic-2+NN	HF	HF/NN
17	1	0.025548	0.054159	0.032338	0.387790	11.99
17	8	0.017975	0.032682	0.020111	0.336192	16.72
17	32	0.017497	0.032367	0.022407	0.337457	15.06
33	1	0.220498	0.492378	0.269099	14.665339	54.50
33	4	0.258773	0.489598	0.266421	29.038271	108.99
33	8	0.639861	1.138679	0.654601	30.930071	47.25

At nodes33 and $B = 4$, HF takes 29.038271 seconds per PDE and the learned evaluator takes 0.266421 seconds, a ratio of 108.9937, or up to 109×. At $B = 8$, these times are 30.930070 and 0.654601 seconds, giving 47.2502×. The larger batch does not improve throughput in this record. It

changes both the batch workload and the execution shape, so the ratio is reported for each measured configuration. This experiment quantifies online response substitution within a fixed composer. Repaying the offline training cost requires a separate workload and accounting of label generation and optimization.

The separate 65-node single-problem study has another timing scope. Its learned setup, solve, and reconstruction-plus-audit phases average 8.23, 17.87, and 104.34 seconds, while FEM setup, solve, and error evaluation average 1.78 seconds. The reference and field-quality comparison are given in Appendix E.3. The batched solve-stage result is therefore not a full-field speedup over sparse FEM.

F.2 A COMPOSED GROUP CAN EXPOSE ANOTHER RELATION

Equations (25) and (26) show that a regular composed group has the same exterior interface type as its components. The earlier implementation materializes a local affine approximation to this exact relation. It stores a reference exterior trace, response, tangent, and the adapter needed to reconstruct internal traces. Independent consistency checks cover response and tangent transformations, save/reload, and evaluation without child-response calls. These checks concern the materialized object and its declared coordinates. Physical accuracy is assessed separately.

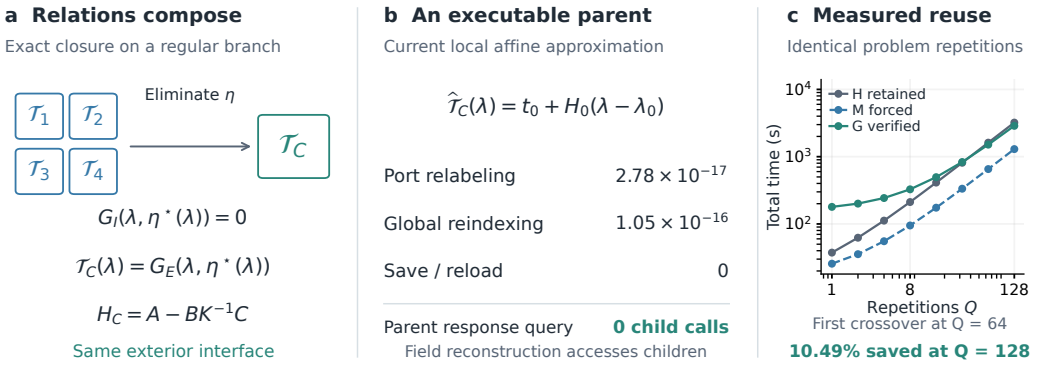

Figure 10: Further reuse of the external relation abstraction. Exact elimination closes the interface type. The current affine materialization is tested as an independently queried object, and a separate workload measures retained-child, parent, and verified selection execution costs. Accuracy verification is paid and tied to the recorded complete-problem input.

The measured one-level workload has 16 rows. The retained-child mode H keeps the children available, mode M uses the affine parent, and mode G selects between them after validity, reference-based accuracy, and cost checks. Before repeated execution, G pays for comparisons to same-port HF and same-mesh full-field FEM. Six parent candidates meet the validity checks, 14 meet the accuracy checks, and five meet all selection conditions. G uses those five parents and retains children for the other 11 rows. The complete output meets the field criterion on 14 of 16 rows. The remaining two also exceed the criterion under retained children.

Each row repeats the same complete problem 128 times with unchanged material, forcing, external boundary data, resolution, and trained weights. The internal trace queries vary during each solve, while the outer problem is repeated. All modes retain prepared contexts. Every repetition executes the chosen solve-and-reconstruct path and recomputes the output. The fixed cost includes context preparation, parent construction, serialization, reload, and reference-based verification. Summing measured phases gives the cost curve. Verified reuse first becomes cheaper at the measured prefix $Q = 64$ and saves 10.49% at $Q = 128$.

This is a verified repeated-execution result. It does not establish a ground-truth-free online accuracy predictor or the cost of final-output memoization, which was not measured. Mathematical closure, independent object consistency, reference-based accuracy, and execution cost answer distinct questions about further reuse.