DISENTANGLING OBSERVATION–SYSTEM COUPLING IN PHYSICAL RECONSTRUCTION VIA ACTIVE SENSING AND INTERVENTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Physical reconstruction for industrial diagnosis, embodied interaction, and physical twins aims to recover object-specific properties that support prediction under new physical inputs. Existing methods often rely on a fixed passive protocol, where the acquisition scheme is predetermined rather than adapted to the observed object. However, distinct physical systems can remain compatible with the same measurements because intrinsic physical effects and observation effects are entangled. In this paper, we propose the **PARAD** (**P**hysics-**A**ware **R**econstruction through **A**ctive **D**isentanglement) framework, which models physical reconstruction through active sensing and intervention and separates physical response generation from measurement formation to analyze such ambiguities. Specifically, PARAD maintains a joint belief over physical and observation parameters from accumulated evidence, and sequentially selects intervention–sensing pairs to distinguish unresolved, task-relevant physical alternatives. Experiments across thermal, vibration, and wave-scattering systems show improved prediction under new inputs, transfer of informative evidence to an unchanged inverse solver. We also propose a foundation for physical reconstruction that actively seeks the evidence needed for prediction, interpretation, and verification.

1 INTRODUCTION

Predicting how individual objects respond to new conditions from limited measurements is a long-standing challenge in industrial diagnosis, embodied interaction, and physical twins (Choi et al., 2008; Feng et al., 2022a; Jiang et al., 2025). Physical reconstruction addresses this challenge by recovering intrinsic properties efficiently enough to support prediction and decision making beyond observed conditions (National Academies of Sciences, Engineering, and Medicine, 2024).

Recent works have advanced physical reconstruction through stronger inverse models (Li et al., 2023; Jiang et al., 2025; Zhong et al., 2026b), active sensing (Jiang et al., 2024; Hemachandra et al., 2025), and intervention design (Shinde et al., 2024; Memmel et al., 2024). Yet accurate reconstruction under limited acquisition remains difficult when physical response and observation effects are entangled. Distinct physical systems may then remain compatible with the same measurements. In the thermal example in Fig. 1, one material difference produces little change under the chosen heating; another produces a distinct response that is obscured by camera blur and sampling. Uncertain sensor parameters can further mask physical differences. This observation–system coupling matters because compatible explanations can predict different responses under new inputs.

The resulting question is what new evidence can distinguish these explanations. Active sensing and Bayesian experimental design provide tools for informative acquisition (Chaloner & Verdinelli, 1995; Hemachandra et al., 2025). Existing methods usually rank experiments by information gain, posterior uncertainty, or task utility, and some already account for nuisance parameters. Their design objective improves inference within a chosen acquisition model, while the physical source of ambiguity remains implicit: a difference may fail to appear under the intervention, or disappear during measurement. PARAD makes these two failure modes explicit and selects intervention–sensing pairs by how well they separate the unresolved physical alternatives that matter to the target response. We build on these tools by distinguishing ambiguity in physical response generation from ambiguity in measurement,

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.

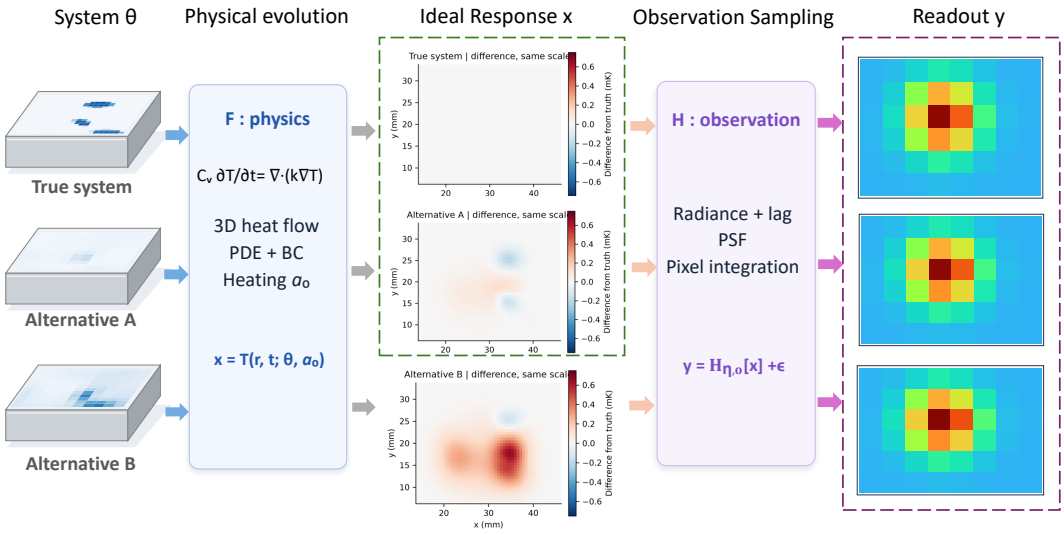


Figure 1: **Two mechanisms of observational equivalence in a thermal schematic.** Rows show the true system and two competing physical hypotheses under the same heating intervention. Alternative A's physical difference is weakly expressed, giving a similar ideal response; Alternative B produces a distinct response, but observation suppresses the difference. Middle panels show temperature differences relative to the true-system response. Similar readouts can therefore leave competing physical explanations unresolved.

and using unresolved, task-relevant physical alternatives to guide intervention–sensing design toward reliable reconstruction and prediction under new conditions.

We propose **PARAD** (**P**hysics-**A**ware **R**econstruction through **A**ctive **D**isentanglement), a modular framework for this process (Fig. 2). Intervention controls which physical differences are expressed; sensing controls which remain measurable. PARAD maintains a joint belief over physical and observation parameters, selects intervention–sensing pairs to distinguish unresolved, task-relevant alternatives, and updates the belief with acquired outcomes. Users instantiate the framework by specifying the physical and observation models, available experimental controls, target properties or responses, and an acquisition budget. Numerical simulators, learned inference modules, and established inverse solvers can implement its interchangeable components, allowing existing reconstruction pipelines to be integrated with subsequent experiment selection rather than replaced.

Across thermal, vibration, and wave-scattering systems, PARAD improves physical reconstruction and prediction under new inputs within limited acquisition budgets, with held-out response error reduced by up to 90.89% in the evaluated benchmarks. Across these settings, its instantiations achieve the strongest reconstruction accuracy or accuracy–acquisition trade-offs among the evaluated methods, while remaining compatible with public simulation benchmarks, real vibration video, and an unchanged external inverse solver. Together, these results support active disentanglement as a general route to obtaining more task-relevant physical information from limited experiments.

This paper proposes three main contributions. First, we introduce an active-disentanglement paradigm for physical reconstruction by separating physical response generation from measurement formation and targeting the unresolved physical distinctions that matter for downstream prediction. Second, we develop a modular framework with interchangeable physical models, inference modules, and acquisition policies, enabling learned components and established reconstruction pipelines to be reused across different physical systems. Third, across thermal, vibration, and scattering systems, we achieve substantial gains in reconstruction and held-out prediction, including up to 90.89% lower response error, while new evidence improves an unchanged inverse solver and excludes observation-confounded physical alternatives.

2 Related Work

2.1 Physical reconstruction and learned inverse models

Visual Vibration Tomography (VVT) recovers interior stiffness and density from monocular vibration observations (Feng et al., 2022a). PAC-NeRF estimates geometry and physical parameters with continuum simulation and neural rendering (Li et al., 2023); particle optimization and progressive joint optimization address errors in geometry and coupled parameter estimation (Kaneko, 2024; Rho et al., 2026). PhysTwin connects reconstruction to simulation under new interactions, while NeFTY applies differentiable heat transport to subsurface thermal reconstruction (Jiang et al., 2025; Zhong et al., 2026b). PARAD complements these methods by selecting evidence that can improve prediction without replacing the inverse solver.

Learned operators and physics-informed models provide complementary inference components (Li et al., 2021; Raissi et al., 2019; Cho & Son, 2025). Diffusion methods combine learned priors with measurement consistency, including DPS and DiffPIR (Chung et al., 2023b; Zhu et al., 2023); physics-aligned diffusion bridges reconstruct fields from sparse observations (Li et al., 2025). BlindDPS jointly treats an unknown imaging operator and image (Chung et al., 2023a). PARAD asks which additional experiments can exclude physical alternatives that fit the existing measurements under admissible observation parameters. InverseBench provides a public setting for evaluating learned priors on scientific inverse problems (Zheng et al., 2025).

2.2 Experimental design through sensing and intervention

Active sensing chooses informative measurements, as in radiance-field view selection and visual–tactile shape reconstruction (Jiang et al., 2024; Smith et al., 2021). PIED optimizes observation locations with physics-informed models for one-shot experimental deployment (Hemachandra et al., 2025). Physical intervention acts earlier in the measurement chain: JIGGLE selects tissue manipulations to estimate attachment boundaries, and ASID learns exploratory actions for system identification (Shinde et al., 2024; Memmel et al., 2024). PARAD treats intervention and sensing as distinct controls for distinguishing task-relevant physical alternatives under persistent observation uncertainty. Its F/H model separates weakly expressed physical differences from differences suppressed or compensated during measurement.

2.3 Uncertainty, sequential inference, and design objectives

Bayesian inverse analysis weights unresolved explanations, while set-membership methods characterize parameters consistent with bounded measurement error (Stuart, 2010; Jaulin & Walter, 1993). Bayesian experimental design supports task-dependent utilities (Chaloner & Verdinelli, 1995), with goal-oriented criteria for prediction (Attia et al., 2018; Zhong et al., 2026a) and formulations incorporating nuisance uncertainty (Bartuska et al., 2022). Variational information estimators, amortized adaptive design, and reinforcement learning make sequential policies computationally accessible (Foster et al., 2019; 2021; Shen et al., 2025); OptBayesExpt implements adaptive measurement selection (McMichael et al., 2021). PARAD builds on these tools and defines disentanglement by task-relevant agreement among measurement-compatible physical explanations. Local nuisance projection and finite-alternative tests assess physical distinguishability, separately from recovery accuracy and prediction under new inputs.

3 PARAD: Active Physical Reconstruction

3.1 Physical Model and Active Disentanglement

A physical experiment maps intrinsic properties to measurements through two stages. The response operator F generates a physical field or trajectory under an intervention, while the observation operator H samples or projects that response into data. Intervention therefore changes which physical differences are expressed; sensing changes which of them remain measurable. We consider fixed material properties and reproducible experiments on systems that can relax or be reset between interventions; Appendix A specifies the operator scope.

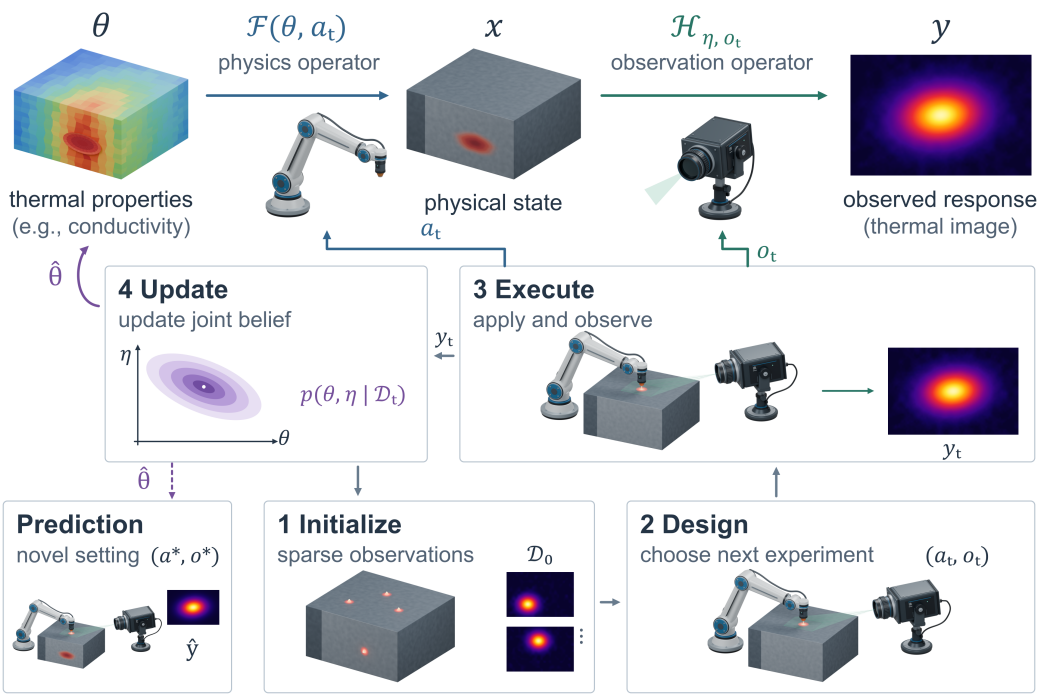


Figure 2: **PARAD: active physical reconstruction.** Intervention a and sensing o control physical response generation F and measurement formation H. Starting from initial observations, PARAD selects an experiment, acquires its outcome, and updates a joint belief over physical parameters θ and observation parameters η. Reconstructed properties are reused to predict responses under new inputs. Existing simulators, learned modules, and inverse solvers can instantiate these interfaces.

Under a fixed protocol, parameter pairs producing exactly the same noiseless readings form an observational fiber, or equivalence class. After choosing a noise model and an explicit measurement tolerance, this equality can be relaxed to a finite-resolution compatibility set. A prior and likelihood define probabilistic weights separately: a Gaussian posterior is not, in general, restricted to this hard set, and its concentration alone does not establish physical identifiability.

Intervention a specifies how the object is excited, sensing o specifies how its response is recorded, x is the ideal physical response, y is the measurement, and ϵ denotes measurement noise. Let θ be intrinsic properties and η persistent observation uncertainty. An unknown that changes physical evolution belongs in the physical model, not solely in η. Let $\mathcal{D}_0$ be the initial acquired record and $\mathcal{D}_t$ its ordered extension by t additional acquisitions, retaining repeated measurements as separate records. The framework is summarized by six relations:

$$\text{Experiment:} \quad x = F(\theta; a), \qquad y = H_{\eta,o}(x) + \epsilon. \tag{1}$$

$$\text{Explanations:} \quad \mathcal{C}_t = \{(\theta, \eta) \text{ consistent with } \mathcal{D}_t\}. \tag{2}$$

$$\text{New evidence:} \quad \mathcal{C}_{t+1} = \mathcal{C}_t \cap S_{t+1}. \tag{3}$$

$$\text{Belief:} \quad q_{t+1}(\theta, \eta) \propto q_t(\theta, \eta) p(y_{t+1} \mid \theta, \eta, a_{t+1}, o_{t+1}, \mathcal{D}_t). \tag{4}$$

$$\text{Planning:} \quad (a_{t+1}, o_{t+1}) = \Pi(q_t, \mathcal{D}_t). \tag{5}$$

$$\text{Local geometry:} \quad J_{\text{eff}} = P_{\text{range}(\widetilde{J}_\eta)^\perp} \widetilde{J}_\theta. \tag{6}$$

Here S_{t+1} contains the explanations consistent with the acquired $(a_{t+1}, o_{t+1}, y_{t+1})$. Equation (3) retains all previous consistency constraints and tolerances; contraction need not be strict. The joint belief q_t denotes the Bayesian target, approximated in computation, and Π is constrained by the available experiments and acquisition budget. Appendix A.1 specifies consistency and joint inference.

Physical and task-level resolution. For a nonempty $\mathcal{C}_t$ on the declared parameter domains, choose a physical target $Q(\theta)$ and a tolerance ε_Q in a prescribed norm. Task-relevant ambiguity is resolved

at this tolerance when

$$\Delta_Q(\mathcal{C}_t) := \sup_{(\theta,\eta),(\theta',\eta')\in\mathcal{C}_t} \|Q(\theta) - Q(\theta')\| \le \varepsilon_Q. \tag{7}$$

Choosing $Q(\theta) = \theta$ tests physical-property resolution; $Q(\theta) = F(\theta; a^\star)$ tests agreement on a response under a new input. Neither requires a unique η. Excluding a fixed physical alternative requires incompatibility for every admissible shared η under the stated tolerance. Our finite-alternative checks do not evaluate the full supremum. Figure 5 in Appendix B illustrates the geometry.

Physical sensitivity after observation compensation. At a differentiable reference $(\bar{\theta}, \bar{\eta})$, let $\bar{x} = F(\bar{\theta}; a)$. The sensitivities are

$$J_\theta = D_x H_{\bar{\eta},o}(\bar{x})\, \partial_\theta F(\bar{\theta}; a), \qquad J_\eta = \partial_\eta H_{\eta,o}(\bar{x})|_{\bar{\eta}}\,. \tag{8}$$

A physical difference can be weakly expressed by $\partial_\theta F$, suppressed by $D_x H$, or compensated by varying η. Stack these sensitivities over $\mathcal{D}_t$ and whiten by its fixed noise covariance. For one unrestricted nuisance perturbation $\delta\eta$ shared across all acquisitions,

$$\min_{\delta\eta} \|\widetilde{J}_\theta \delta\theta + \widetilde{J}_\eta \delta\eta\|_2 = \|J_{\text{eff}}\delta\theta\|_2. \tag{9}$$

The right-hand side is the physical sensitivity that no shared observation adjustment can cancel to first order. With known calibration, the nuisance space is absent and $J_{\text{eff}} = \widetilde{J}_\theta$. This first-order diagnostic does not establish global identifiability (Appendix A.2).

Limits of one-sided acquisition. Equal ideal responses under fixed excitation remain indistinguishable for every sensing choice at the same η. With fixed sensing, the noiseless response $\mu(a) = \eta a/\theta$ for $\theta, \eta > 0$ is invariant under admissible rescaling $(\theta, \eta) \mapsto (c\theta, c\eta)$, $c > 0$, $c \neq 1$, regardless of a. Under equal noise laws, these ambiguities also survive repetition (Appendix A.3). Joint access enlarges the experiment family; feedback selects within it, without guaranteeing improvement over an informative fixed sequence.

3.2 Using PARAD for a Reconstruction Task

To instantiate PARAD, specify the physical task and connect the following interfaces.

(1) Define the physical target. Choose a representation θ for the unknown object properties and a target $Q(\theta)$, such as a material field or a response under a specified new input. This target determines which unresolved physical differences the acquisition policy should distinguish.

(2) Separate response from measurement. Specify $F(\theta; a)$, $H_{\eta,o}$, and the measurement-noise model. For thermal reconstruction, F predicts temperature, a controls heating, and $H_{\eta,o}$ describes camera observation under sensing settings o. A sensing operation that perturbs the physical system also has an intervention component.

(3) Specify evidence and admissible experiments. Record initial measurements together with their applied intervention–sensing pairs. Define admissible experiments, acquisition costs, and the total budget. Declare the prior, supplied calibration, and pretrained assets available to inference and planning. The object's unknown physical properties and unacquired response values are not available to these modules.

(4) Attach inference and planning modules. Use the full history to update the physical estimate and an approximation to q_t. Known calibration fixes η; otherwise one unknown η is shared across acquisitions. A point-estimate inverse solver can be paired with a separate planning approximation. Local sensitivities or candidate predictions can score experiments according to Q.

(5) Acquire, update, and reuse. Select an admissible experiment using the current belief, execute it, and append its measured outcome to the history. The acquired outcome, rather than its forecast, enters Equation (4). Repeat within the acquisition budget, then reuse the saved physical estimate to predict new responses without refitting to their measurements.

Our implementations use pretrained thermal networks, a retained VVT inverse core, and a frozen scattering diffusion prior (Appendices C–E). The evaluated feedback policies are model-based; Appendix A.4 gives a local design realization and its computational interfaces.

4 EXPERIMENTS

We test physical reuse, compatibility with external reconstruction modules, and exclusion of observation-confounded physical alternatives. The systems span heat diffusion, elastic vibration, and wave scattering, using a generated benchmark, public simulations, and real vibration video. Within each comparison, acquisition budgets and evaluation rules are matched; the acquired records need not be identical. Priors, representations, and computation are specified separately. All response predictions use frozen material or field estimates.

4.1 THERMAL RECONSTRUCTION: WHAT NEW MEASUREMENTS ADD

Physical task and controls. Subsurface resistance changes heat transport while the camera observes its surface signature (Choi et al., 2008; Zhong et al., 2026b). Our 3D benchmark reconstructs a 32×32 resistance field from 16×16 coefficients and a continuous depth, jointly with three shared emissivity coordinates. Heating defines a and camera sampling defines o. We evaluate four newly generated instances of the same thermal benchmark. They share a three-acquisition prefix; the fourth acquisition follows one of three protocol rules: Fixed (TRAIN-designed), Universal fixed (DEV-selected), or Feedback (selected from the current object's history among 13 joint protocols). All spend $K = 4$ acquisitions (0.16 J), sharing the frozen networks, initialization, prior, and physical-refinement endpoint. This isolates one adaptive transition (Appendix C).

Table 1: Thermal reconstruction on four new TEST objects. All rows use four paid acquisitions (0.16 J). The first two rows compare inference on identical histories; the last three compare acquisition at the same physical endpoint. Mean errors over the four TEST objects; lower is better.

Acquisition	Reconstruction	Field	Depth (mm)	New task NRMSE
Fixed	Original	0.1664	0.1886	0.4608
Fixed	Physical refinement	**0.0620**	0.0249	0.0924
Universal fixed	Physical refinement	**0.0620**	0.0240	0.0762
Feedback	Physical refinement	0.0661	**0.0205**	**0.0668**

Physical refinement of learned inference. On identical Fixed histories, numerical F/H refinement reduces field error from 0.1664 to 0.0620 and policy-unused response NRMSE from 0.4608 to 0.0924. Both estimates share frozen-network initialization and surrogate optimization; the comparison isolates the added numerical F/H refinement. The three response tasks contain two new heating loads and one unmeasured internal response under an acquired load (Fig. 3a). These tasks are excluded from the planner utility, although their DEV results were available during development. Target-task definitions and TRAIN-object responses were used for task-model training; the held-out quantities are the TEST object's unmeasured responses.

What changing the experiment adds. At the same refinement endpoint, Universal fixed and feedback increase nuisance-profiled separation of fixed physical alternatives by 24.22 and 24.02, versus 8.15 for Fixed. The tested direction exhibits limited nuisance compensation, so these gains primarily reflect improved observability. Feedback further lowers new-task error from 0.0762 to 0.0668 relative to Universal fixed, with slightly higher field error. Together, the four-object comparison separates improved observability from task-specific prediction gains at a matched acquisition budget.

4.2 VIBRATION TOMOGRAPHY: CHANGING EVIDENCE WITHOUT CHANGING THE SOLVER

Public simulation and complete-system comparison. Six public COMSOL inclusion objects from VVT (Feng et al., 2022a) provide x/z motion at 25 surface points. Both systems purchase back release and two further 3-second packets from left, right, and twist. VVT-adapted retains the official inverse core with a motion-to-mode frontend; ours uses a nine-parameter inclusion model and two history-dependent selections. Acquisition budgets are matched, but the selected records, representation, mesh, and inference can differ. All six objects have historical development exposure. Material log-RMSE and unacquired front-response error decrease by 67.38% and 90.89%,

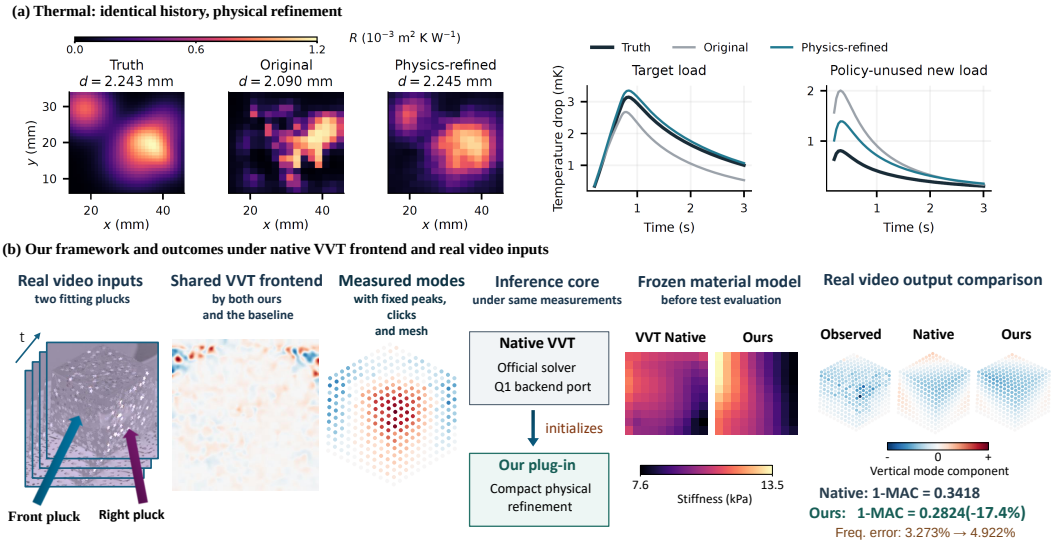


Figure 3: Physical reuse under frozen estimates. (a) The first prespecified thermal TEST object: same-history material estimates and a policy-unused response. (b) Real AVI frames, shared VVT preprocessing, and VVT-initialized refinement lead to frozen materials and held-out modal predictions. The specimen has a documented clay inclusion (Feng et al., 2022a); material slices are estimates, not internal ground truth. Response panels show the sole mode absent from fitting videos, with display-only sign/amplitude alignment; numbers average all three evaluated modes (Table 2C). Held-out dynamics do not enter fitting; static calibration and fixed extraction bins remain known.

respectively (Table 2A). Front motion is excluded from fitting and planning, then scored over all 6,000 frames, 1,331 nodes, and three components without amplitude/phase alignment.

Table 2: Vibration evidence at three levels. Every block uses the same left-to-right convention: reference versus comparison. A: complete systems on six public simulation objects. B: evidence replay through the same VVT-adapted inverse core. C: one real object with two fitting videos and one held-out video; means over three modes. Errors ↓.

Metric / case	Reference	Comparison
A. Complete systems (simulation)	**VVT-adapted**	**Ours**
Material R	1.12037	**0.36544**
Unacquired front P	0.85848	**0.07825**
B. Same VVT core	**Fixed history**	**Selected history**
04: front P; different actions	0.49066	**0.38468**
07: front P; different actions	1.64641	**0.34668**
C. Raw real video	**Native VVT**	**VVT + refinement**
Frequency error (%)	**3.273**	4.922
Shape error $1 - \mathrm{MAC}$	0.3418	**0.2824**

A matches acquisition budgets, not necessarily selected records, representation, or inference. B keeps the VVT-adapted inverse implementation fixed and changes only the acquired evidence. C shares the official frontend and measurements; internal material ground truth is unavailable. Native VVT uses the official inverse core with a verified Windows Q1 assembly backend rather than a full DOLFIN-environment replay.

Changing evidence while retaining the inverse core. Passing our acquired history to the unchanged VVT-adapted pipeline reduces front-response error from 0.49066 to 0.38468 on object 04 and from 1.64641 to 0.34668 on object 07 (Table 2B). As replay controls, preserving the same packet set reproduces the original endpoint: object 05 remains at $P = 0.77232$, while object 10 remains at

$P = 0.39539$ even when the left/right packet order is reversed. These zero-change controls verify that replay itself does not introduce spurious gains, supporting attribution of the 04/07 improvements to the substituted evidence under the same VVT core.

From raw video to held-out modal prediction. Three official AVI recordings depict one gelatin specimen with a documented clay inclusion (Feng et al., 2022a). Front and right plucks fit its materials; left-side-twist dynamics are withheld. Figure 3b shows the raw-input path through shared VVT preprocessing to frozen-material prediction. Both methods share clicks, fixed peaks, geometry, and a 10^3 Q1 mesh. Our 16-coefficient refinement starts from the official VVT solution; checkpoint selection uses only the fitting objective. Re-solving with frozen materials reduces mean held-out shape error $1 - \mathrm{MAC}$ from 0.3418 to 0.2824 (17.4%), while frequency error rises from 3.273% to 4.922%. Only one of the three evaluated modes is absent from the fitting videos; static test-view calibration remains known. The real-video experiment connects the shared measurement frontend to material refinement and subsequent modal prediction with frozen properties.

4.3 Wave-scattering tomography: Better Recovery Does Not Remove Ambiguity

Controlled high-dimensional recovery. InverseBench supplies 128×128 fields and a linear scattering operator (Zheng et al., 2025). We add five shared unknown receiver gain/phase parameters. The new comparison fixes four of 20 transmitters and all 360 receivers. DPS, RED-diff, and our refinements share acquired data and the same frozen diffusion checkpoint (Chung et al., 2023b; Mardani et al., 2024). R_0 is 1,000-step DPS plus physical refinement; R_s adds five denoising/physics stages; R_+ uses four-rotation-averaged denoising at those stages and a final acquired-data repair. No weights are trained and no observations added. All five methods were frozen before evaluating 24 previously unexamined confirmation objects under three receiver conditions (360/360 runs); pretraining-set overlap remains unverified.

Table 3: New fixed-$K = 4$ scattering confirmation: 24 objects, three H conditions, common frozen prior and blind-H interface. Means first average conditions within objects. C_H: cross-H field variation; time includes the complete recovery chain. Best means in this tested comparison are bold.

Recovery	PSNR ↑	SSIM ↑	$E_\theta \downarrow$	$Q_F \downarrow$	$C_H \downarrow$	s ↓
DPS	22.287	0.7435	0.2708	0.08276	0.11390	35.12
R_0: DPS + physics	22.457	0.8318	0.2665	0.07403	0.10779	41.67
R_s: staged prior/physics	23.316	0.8358	0.2413	0.06708	0.06157	48.13
RED-diff	21.061	0.8185	0.3079	0.08911	**0.02837**	**21.79**
R_+: averaged prior/repair	**24.156**	**0.8522**	**0.2160**	**0.05963**	0.05226	54.04

A stronger recovery module with unchanged learned weights. Relative to R_0, R_+ improves PSNR by 1.70 dB, SSIM by 0.0205, field error by 18.95%, and unmeasured-transmitter error Q_F by 19.44%. Each improves on 23/24 object means. The paired 95% interval for ΔQ_F is $[-0.01698, -0.01141]$. Since Q_F evaluates F without H, this is physical prediction rather than receiver recalibration alone. Full runtime rises from 41.67 to 54.04 s; RED-diff remains faster and has lower cross-H field variation. These gains demonstrate the effectiveness of the complete staged prior–physics update with a shared frozen checkpoint.

Distinct physical explanations remain compatible after refinement. A post-hoc check finds both saved R_0/R_+ fields compatible with the same four acquired records at the same true H in all 72 object/condition pairs. Their mean symmetric field distance is 0.159, and their unmeasured F predictions also differ. The two refinements therefore select different physical explanations from the same measurement-compatible set, separating recovery quality from the information supplied by acquisition. On the eight-object known-H subset, R_+ still improves all four recovery metrics over R_0 on every object (Appendix E).

What additional evidence can exclude. A separate historical DEV analysis fixes twelve primary alternatives on four objects. Seven across three objects fail the initial-data tolerance at true H but fit under a legal shared H. After one additional record, even a relaxed residual lower bound exceeds

tolerance, excluding those fixed fields for every legal H (Fig. 4). Both the tested added acquisition and strong fixed controls exclude these alternatives. The test quantifies how new evidence resolves specific physical disagreements sustained by observation-side compensation.

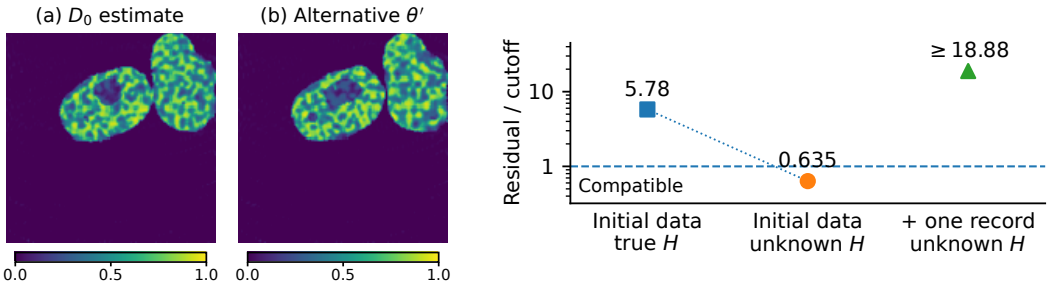


Figure 4: Historical DEV mechanism test, separate from the new 24-object comparison. A fixed alternative fails at true H, fits the initial data under legal unknown H, and is excluded by a relaxed lower bound after one new record. Values are normalized by stage-specific noise-resolution cutoffs; compatibility lies below 1. The last value is a lower bound for every legal H. The displayed fields are an estimate and an alternative, not truth and reconstruction.

5 Conclusion

We formulate physical reconstruction as active disentanglement: experiments are used to distinguish task-relevant physical explanations rather than only refine estimates on fixed data. Across thermal, vibration, and scattering systems, PARAD connects heterogeneous reconstruction modules with active acquisition while separating recovery gains from information gains supplied by new evidence. This view makes observation–system coupling an explicit target for physical reconstruction under limited acquisition.

Limitations: Our experiments establish the value of changing the acquired evidence, but do not establish a all-round advantage of feedback over a sufficiently informative fixed design. In addition, our framework in this version relies on a selected physical representation θ, forward model F, and application-specific components. At theoretical level, PARAD provides an explicable framework for active disentanglement based on a simplified physical model instead of a general proof of the global-identifiability.

Future works: Future work should move upstream from experiment selection toward learning or discovering the physical representation θ, forward model F, and sensing interfaces, reducing application-specific specification. It should also study more sample-efficient learned planners and real closed-loop systems under unknown or time-varying observation processes. On the theory side, extending the current local and finite-alternative analysis toward broader nonlinear identifiability and principled stopping criteria is an important next step.

AI Use Statement

AI assistants supported coding, manuscript drafting, and illustrative figure preparation. The authors developed the core ideas and theoretical analysis, interpreted the results, and reviewed all AI-assisted content. Quantitative figures and tables were generated from saved experimental outputs. Experimental procedures and data provenance are documented in the appendix.

Ethics Statement

The study uses generated physical simulations and publicly available vibration recordings of a gelatin specimen. Safety-critical deployment, including structural diagnosis, requires independent validation and application-specific uncertainty assessment.

Reproducibility Statement

Section 3 defines the model and its inference/design interfaces. Appendix A provides the theoretical assumptions, proofs, and computational details, and Appendix B gives the analytical beam example. Appendices C–E specify the experimental protocols, data permissions, training or checkpoint dependencies, evaluation rules, and computational settings.

References

Ahmed Attia, Alen Alexanderian, and Arvind K. Saibaba. Goal-oriented optimal design of experiments for large-scale Bayesian linear inverse problems. *Inverse Problems*, 34(9):095009, 2018. doi: 10.1088/1361-6420/aad210. URL https://doi.org/10.1088/1361-6420/aad210.

Arved Bartuska, Luis Espath, and Raúl Tempone. Small-noise approximation for Bayesian optimal experimental design with nuisance uncertainty. *Computer Methods in Applied Mechanics and Engineering*, 399:115320, 2022. doi: 10.1016/j.cma.2022.115320. URL https://doi.org/10.1016/j.cma.2022.115320.

Kathryn Chaloner and Isabella Verdinelli. Bayesian experimental design: A review. *Statistical Science*, 10(3):273–304, 1995. doi: 10.1214/ss/1177009939. URL https://doi.org/10.1214/ss/1177009939.

Sung Woong Cho and Hwijae Son. Physics-informed deep inverse operator networks for solving PDE inverse problems. In *International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=0FxnSZJPmh.

Manyong Choi, Kisoo Kang, Jeonghak Park, Wontae Kim, and Koungsuk Kim. Quantitative determination of a subsurface defect of reference specimen by lock-in infrared thermography. *NDT & E International*, 41(2):119–124, 2008. doi: 10.1016/j.ndteint.2007.08.006. URL https://www.sciencedirect.com/science/article/pii/S096386950700103X.

Hyungjin Chung, Jeongsol Kim, Sehui Kim, and Jong Chul Ye. Parallel diffusion models of operator and image for blind inverse problems. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6059–6069, 2023a. doi: 10.1109/CVPR52729.2023.00587. URL https://doi.org/10.1109/CVPR52729.2023.00587.

Hyungjin Chung, Jeongsol Kim, Michael Thompson McCann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *International Conference on Learning Representations*, 2023b. URL https://openreview.net/forum?id=OnD9zGAGT0k.

Berthy T. Feng, Alexander C. Ogren, Chiara Daraio, and Katherine L. Bouman. Visual vibration tomography: Estimating interior material properties from monocular video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16231–16240, 2022a. URL https://openaccess.thecvf.com/content/CVPR2022/html/Feng_Visual_Vibration_Tomography_Estimating_Interior_Material_Properties_From_Monocular_Video_CVPR_2022_paper.html.

Berthy T. Feng, Alexander C. Ogren, Chiara Daraio, and Katherine L. Bouman. Visual vibration tomography: Official implementation. Software repository, 2022b. URL https://github.com/berthyf96/visvib-tomography. Experimental baseline pinned to commit 4757f8298b2989c1b647eba818daf7926aef4ab4.

Adam Foster, Martin Jankowiak, Elias Bingham, Paul Horsfall, Yee Whye Teh, Thomas Rainforth, and Noah Goodman. Variational Bayesian optimal experimental design. In *Advances in Neural Information Processing Systems*, volume 32, pp. 14036–14047, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/d55cbf210f175f4a37916eafe6c04f0d-Abstract.html.

Adam Foster, Desi R. Ivanova, Ilyas Malik, and Tom Rainforth. Deep adaptive design: Amortizing sequential Bayesian experimental design. In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 3384–3395. PMLR, 2021. URL https://proceedings.mlr.press/v139/foster21a.html.

Apivich Hemachandra, Gregory Kang Ruey Lau, See-Kiong Ng, and Bryan Kian Hsiang Low. PIED: Physics-informed experimental design for inverse problems. In *International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=w7P92BEsb2.

Luc Jaulin and Eric Walter. Set inversion via interval analysis for nonlinear bounded-error estimation. *Automatica*, 29(4):1053–1064, 1993. doi: 10.1016/0005-1098(93)90106-4. URL https://doi.org/10.1016/0005-1098(93)90106-4.

Hanxiao Jiang, Hao-Yu Hsu, Kaifeng Zhang, Hsin-Ni Yu, Shenlong Wang, and Yunzhu Li. PhysTwin: Physics-informed reconstruction and simulation of deformable objects from videos. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 7219–7230, 2025. URL https://openaccess.thecvf.com/content/ICCV2025/html/Jiang_PhysTwin_Physics-Informed_Reconstruction_and_Simulation_of_Deformable_Objects_from_Videos_ICCV_2025_paper.html.

Wen Jiang, Boshu Lei, and Kostas Daniilidis. FisherRF: Active view selection and mapping with radiance fields using fisher information. In *Computer Vision – ECCV 2024*, volume 15071 of *Lecture Notes in Computer Science*, pp. 422–440. Springer Nature Switzerland, 2024. doi: 10.1007/978-3-031-72624-8_24. URL https://doi.org/10.1007/978-3-031-72624-8_24.

Takuhiro Kaneko. Improving physics-augmented continuum neural radiance field-based geometry-agnostic system identification with Lagrangian particle optimization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5470–5480, 2024. doi: 10.1109/CVPR52733.2024.00523. URL https://doi.org/10.1109/CVPR52733.2024.00523.

Xuan Li, Yi-Ling Qiao, Peter Yichen Chen, Krishna Murthy Jatavallabhula, Ming Lin, Chenfanfu Jiang, and Chuang Gan. PAC-NeRF: Physics augmented continuum neural radiance fields for geometry-agnostic system identification. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=tVkrbkz42vc.

Zeyu Li, Hongkun Dou, Shen Fang, Wang Han, Yue Deng, and Lijun Yang. Physics-aligned field reconstruction with diffusion bridge. In *International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=D042vFwJAM.

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

Morteza Mardani, Jiaming Song, Jan Kautz, and Arash Vahdat. A variational perspective on solving inverse problems with diffusion models. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=1YO4EE3SPB.

Robert D. McMichael, Sean M. Blakley, and Sergey Dushenko. Optbayesexpt: Sequential Bayesian experiment design for adaptive measurements. *Journal of Research of the National Institute of Standards and Technology*, 126:126002, 2021. doi: 10.6028/jres.126.002. URL https://doi.org/10.6028/jres.126.002.

Marius Memmel, Andrew Wagenmaker, Chuning Zhu, Dieter Fox, and Abhishek Gupta. ASID: Active exploration for system identification in robotic manipulation. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/ae8afec5dabb07091d8f85128250f2ff-Abstract-Conference.html.

National Academies of Sciences, Engineering, and Medicine. *Foundational Research Gaps and Future Directions for Digital Twins*. The National Academies Press, Washington, DC, 2024. doi: 10.17226/26894. URL https://doi.org/10.17226/26894.

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. URL https://doi.org/10.1016/j.jcp.2018.10.045.

Daniel Rho, Jun Myeong Choi, Biswadip Dey, and Roni Sengupta. ProJo4D: Progressive joint optimization for sparse-view inverse physics estimation. *Transactions on Machine Learning Research*, 2026. URL https://openreview.net/forum?id=pqvVrqlXCZ.

Wanggang Shen, Jiayuan Dong, and Xun Huan. Variational sequential optimal experimental design using reinforcement learning. *Computer Methods in Applied Mechanics and Engineering*, 444: 118068, 2025. doi: 10.1016/j.cma.2025.118068. URL https://doi.org/10.1016/j.cma.2025.118068.

Nikhil Uday Shinde, Xiao Liang, Fei Liu, Yutong Zhang, Florian Richter, Sylvia Lee Herbert, and Michael C. Yip. JIGGLE: An active sensing framework for boundary parameters estimation in deformable surgical environments. In *Proceedings of Robotics: Science and Systems*, Delft, Netherlands, 2024. doi: 10.15607/RSS.2024.XX.007. URL https://www.roboticsproceedings.org/rss20/p007.html.

Edward Smith, David Meger, Luis Pineda, Roberto Calandra, Jitendra Malik, Adriana Romero Soriano, and Michal Drozdzal. Active 3D shape reconstruction from vision and touch. In *Advances in Neural Information Processing Systems*, volume 34, pp. 16064–16078, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/8635b5fd6bc675033fb72e8a3ccc10a0-Abstract.html.

Andrew M. Stuart. Inverse problems: A Bayesian perspective. *Acta Numerica*, 19:451–559, 2010. doi: 10.1017/S0962492910000061. URL https://doi.org/10.1017/S0962492910000061.

Hongkai Zheng, Wenda Chu, Bingliang Zhang, Zihui Wu, Austin Wang, Berthy T. Feng, Caifeng Zou, Yu Sun, Nikola Kovachki, Zachary E. Ross, Katherine L. Bouman, and Yisong Yue. InverseBench: Benchmarking plug-and-play diffusion priors for inverse problems in physical sciences. In *International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=U3PBITXNG6.

Shijie Zhong, Wanggang Shen, Tommie Catanach, and Xun Huan. Goal-oriented Bayesian optimal experimental design for nonlinear models using Markov chain Monte Carlo. *SIAM/ASA Journal on Uncertainty Quantification*, 14(1):19–47, 2026a. doi: 10.1137/24M1649344. URL https://doi.org/10.1137/24M1649344.

Tao Zhong, Yixun Hu, Dongzhe Zheng, Aditya Sood, and Christine Allen-Blanchette. Neural field thermal tomography: A differentiable physics framework for non-destructive evaluation. *arXiv preprint arXiv:2603.11045*, 2026b. URL https://arxiv.org/abs/2603.11045.

Yuanzhi Zhu, Kai Zhang, Jingyun Liang, Jiezhang Cao, Bihan Wen, Radu Timofte, and Luc Van Gool. Denoising diffusion models for plug-and-play image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1219–1229, 2023. URL https://openaccess.thecvf.com/content/CVPR2023W/NTIRE/html/Zhu_Denoising_Diffusion_Models_for_Plug-and-Play_Image_Restoration_CVPRW_2023_paper.html.

A Framework Details and Theoretical Analysis

Operator scope. F is a solution operator returning an ideal field or a reproducible response trajectory. It can be evaluated analytically, by a numerical solver, or by a learned operator (Li et al., 2021). We consider fixed material properties and reproducible experiments on systems that can relax or be reset between interventions. The intervention a changes prescribed forcing, initial conditions, or boundary conditions. The sensing protocol o changes how the resulting response is sampled, projected, or integrated by $H_{\eta,o}$; η represents persistent observation-side uncertainty. An unknown that changes physical evolution belongs in the physical model, rather than solely in η. A sensing operation that also perturbs the physics has a component in a. Visual encoders can supply observation preprocessing or inference representations.

A.1 Exact Equivalence, Noisy Consistency, and Joint Inference

What the evidence distinguishes. Fix admissible parameter domains Θ and $\mathcal{E}$. For an ordered protocol history $\mathcal{P} = ((a_i, o_i))_{i=1}^n$, define $M_{\mathcal{P}}(\theta, \eta) = [H_{\eta,o_i}(F(\theta; a_i))]_{i=1}^n$. Noiseless equal-response fibers are equivalence classes. Global physical decoupling on $\Theta \times \mathcal{E}$ means that, for all admissible pairs,

$$M_{\mathcal{P}}(\theta, \eta) = M_{\mathcal{P}}(\theta', \eta') \implies \theta = \theta'. \tag{10}$$

This does not require a unique η. For a nonempty noisy compatibility set, the physical-resolution criterion is $\sup_{(\theta,\eta),(\theta',\eta')\in\mathcal{C}_t} \|\theta - \theta'\| \le \varepsilon_\theta$ in a prescribed physical norm. A task-specific criterion instead bounds $\|Q(\theta) - Q(\theta')\|$ over the same pairs. Resolving one observed fiber is weaker than establishing Equation (10) over the whole domain. These are definitions of distinguishability.

A nested consistency construction. For a known positive-definite noise covariance R_o independent of the unknown parameters, fix a tolerance $\tau_{a,o} \ge 0$ before fitting and define

$$S(a, o, y) = \left\{(\theta, \eta) \in \Theta \times \mathcal{E} : \left\| R_o^{-1/2}\left[y - H_{\eta,o}(F(\theta; a))\right]\right\|_2 \le \tau_{a,o}\right\}. \tag{11}$$

Intersecting these sets over all records in $\mathcal{D}_t$, including $\mathcal{D}_0$, defines $\mathcal{C}_t$ in Equation (2). With $S_{t+1} = S(a_{t+1}, o_{t+1}, y_{t+1})$, retaining the earlier constraints gives Equation (3). The sets can be empty or disconnected and are not automatically calibrated confidence regions. A fresh pooled-residual test with a history-dependent cutoff need not generate nested sets. The scattering checks in Appendix E.4 use such stagewise noise-resolution cutoffs: they establish the stated compatibility or exclusion of tested fields, not monotone contraction of the entire feasible domain.

Joint inference from the complete record. A prior $p_0(\theta, \eta)$ and likelihood define probabilistic weights separately (Stuart, 2010). Under conditionally independent Gaussian acquisition errors with known covariances,

$$q_t(\theta, \eta) = p(\theta, \eta \mid \mathcal{D}_t) \propto p_0(\theta, \eta) \prod_{(a_i,o_i,y_i)\in\mathcal{D}_t} \mathcal{N}(y_i; H_{\eta,o_i}(F(\theta; a_i)), R_{o_i}). \tag{12}$$

The product runs over the ordered records with multiplicity, including the initial record; q_0 therefore already conditions on $\mathcal{D}_0$. The update index t and the total paid acquisition count differ: $K = |\mathcal{D}_0| + t$. Full-history refitting uses p_0 once, whereas Equation (4) adds only the next likelihood to the current belief. A Gaussian likelihood does not restrict inference to Equation (11), and its credible sets need not be nested. Numerical distributions, local approximations, and candidate-based summaries need not realize this Bayesian target exactly. Inference modules approximate this joint Bayesian target using local distributions or candidate-based summaries.

Correlations within an acquisition can be represented by R_o. Equation (4) also allows history-dependent likelihoods; cross-acquisition correlations require these conditional likelihoods rather than the independent product in Equation (12). For independent circular complex Gaussian measurement errors with $\mathbb{E}|\epsilon|^2 = \sigma_y^2$, stacking real and imaginary parts gives covariance $(\sigma_y^2/2)I$ and negative log-likelihood $\|y - \mu\|^2/\sigma_y^2$ up to constants. Thus the real-Gaussian factor $1/2$ and the scattering normalization are consistent under this noise convention. Known calibration fixes η; otherwise the same unknown η is shared across acquisitions.

With adaptive acquisition, the joint history law also contains policy factors $\pi_i(a_i, o_i \mid \mathcal{D}_{i-1})$. When the policy uses only observed history, declared assets, and randomness independent of the unknown object, these factors are identical across hypotheses conditional on the realized history and cancel from posterior odds. This is the non-anticipating condition underlying Equation (4); hidden object truth is not a planner input.

A.2 Nuisance Projection and the Schur Complement

At a reference estimate $(\bar{\theta}, \bar{\eta})$, assume differentiability in the unknown parameters and write $\bar{x} = F(\bar{\theta}; a)$. The predicted mean $\mu = H_{\eta,o}(F(\theta; a))$ has first-order perturbation

$$\begin{aligned} &\delta\mu \simeq J_\theta \delta\theta + J_\eta \delta\eta, \\ &J_\theta = D_x H_{\bar{\eta},o}(\bar{x})\, \partial_\theta F(\bar{\theta}; a), \qquad J_\eta = \partial_\eta H_{\eta,o}(\bar{x})|_{\bar{\eta}}\,. \end{aligned} \tag{13}$$

Stack the Jacobians over the same acquired history and whiten by its fixed observation covariance $\mathcal{R}_t$; for independent acquisitions, $\mathcal{R}_t$ is the block diagonal of their R_{o_i}. Tildes denote these whitened matrices. Complex data are first stacked into real and imaginary parts. One shared nuisance perturbation acts across the entire history. With $I_{uv} = \widetilde{J}_u^\top \widetilde{J}_v$, Equation (6) gives

$$J_{\text{eff}}^\top J_{\text{eff}} = I_{\theta\theta} - I_{\theta\eta} I_{\eta\eta}^\dagger I_{\eta\theta}, \tag{14}$$

where † denotes the Moore–Penrose pseudoinverse.

Proposition 1 (First-order compensation and separation)**.** *For the stacked, whitened sensitivities above and unrestricted linearized nuisance perturbations,*

$$\min_{\delta\eta} \|\widetilde{J}_\theta \delta\theta + \widetilde{J}_\eta \delta\eta\|_2 = \|J_{\text{eff}} \delta\theta\|_2 \geq \sigma_{\min}(J_{\text{eff}}) \|\delta\theta\|_2. \tag{15}$$

Here $\sigma_{\min}(J_{\text{eff}}) = \inf_{\|v\|_2=1} \|J_{\text{eff}} v\|_2$ *in the stated physical coordinates; it is zero for any nontrivial kernel. A physical direction admits exact first-order nuisance compensation if and only if it lies in* $\ker J_{\text{eff}}$.

Proof. Decompose $\widetilde{J}_\theta \delta\theta$ orthogonally into the range of $\widetilde{J}_\eta$ and its complement. A nuisance perturbation can cancel the first component but cannot change the second. This gives the equality; the singular-value bound follows by definition. Moreover, $P_{\text{range}(\widetilde{J}_\eta)} = \widetilde{J}_\eta (\widetilde{J}_\eta^\top \widetilde{J}_\eta)^\dagger \widetilde{J}_\eta^\top$, which yields Equation (14), including for rank-deficient nuisance blocks. □

Projecting each acquisition separately would allow a different nuisance compensation in every round, contrary to the shared-η model. With known η, the nuisance space is absent and $J_{\text{eff}} = \widetilde{J}_\theta$. The projection is a measurement-only diagnostic: prior regularization does not remove a data equivalence. A compensating perturbation may violate calibration bounds, and linearized cancellation need not extend to a feasible nonlinear alternative. Conversely, a zero derivative alone does not prove global non-identifiability. Finite-alternative tests enforce or explicitly relax the admissible nuisance domain separately.

A.3 Limits of One-Sided Access and Fixed Protocols

Proposition 2 (One-sided counterexamples)**.** *For each of observation-only, intervention-only, and repeated fixed-protocol acquisition, there exist models and distinct admissible physical parameters with identical history distributions, including under adaptive choices within the allowed family.*

Proof. For observation-only acquisition, equal ideal responses under the fixed intervention remain equal under every sensing protocol at the same η. For example, let $F(\theta; a) = a_1\theta_1 + a_2\theta_2$ and fix $a = (1, 0)$. Distinct θ_2 values with the same θ_1 then remain indistinguishable under any $H_{\eta,o}$. For intervention-only acquisition at fixed o_0, consider positive θ, η and

$$F(\theta; a) = a/\theta, \qquad H_{\eta,o_0}(x) = \eta x, \qquad \eta a/\theta = (c\eta)a/(c\theta). \tag{16}$$

Every intervention preserves the ambiguity whenever $c > 0$, $c \neq 1$, and both parameter pairs are admissible. With equal noise laws, the conditional measurement distributions coincide. A history-based policy therefore has the same next-design distribution under both hypotheses; induction gives

identical complete history laws. Repetition of a fixed protocol in either example preserves the same ambiguity. □

A fixed sequence varying both controls can succeed when its combined measurements separate the relevant alternatives. Joint access enlarges the admissible experiment family; feedback chooses within it using acquired evidence. Neither guarantees that a particular planner outperforms an informative fixed sequence.

A.4 Computational Modules and Experiment Selection

Inference modules. Thermal reconstruction uses two pretrained networks: `HistoryInverse` initializes (θ, η), and `ForwardSurrogate` supports a differentiable surrogate fit before numerical F/H refinement. Vibration uses classical material fitting and the official VVT inverse core. Scattering uses an external frozen diffusion denoiser within DPS and repeated prior–physics updates. Network weights remain fixed during object-specific inference. Appendices C–E specify these implementations. These implementations demonstrate compatibility with different reconstruction modules.

A local design realization. For a local linear-Gaussian approximation, let z collect the chosen inference coordinates for (θ, η) and let $\Sigma_t \succ 0$ be their joint covariance. This is a parameter covariance, distinct from the observation covariance $\mathcal{R}_t$ in Appendix A.2. For a candidate $p = (a, o)$ with noise independent of the acquired history, let J_p be the noise-whitened measurement Jacobian with respect to all coordinates in z, and let $G_t = \partial_z Q(\theta(z))$ be the Jacobian of the chosen physical target. Target units or scales are specified before scoring. Evaluate both Jacobians at the current estimate. One local utility is

$$\Sigma_t^+(p) = (\Sigma_t^{-1} + J_p^\top J_p)^{-1}, \qquad U_t(p) = \operatorname{tr}\big[G_t(\Sigma_t - \Sigma_t^+(p))G_t^\top\big]. \tag{17}$$

This predicts a nonnegative reduction of target covariance under the local model, not an observed improvement or a global identification guarantee. Although Q has no direct nuisance dependence, physical–nuisance cross-covariances affect its uncertainty. The planner scores admissible protocols within the remaining budget; the acquired outcome, not its prediction, enters Equation (4).

The evaluated feedback policies are model-based, not reinforcement-learned. Thermal design uses task-response sensitivity; vibration uses scale-projected material sensitivity. Historical scattering policies use candidate fields and field- or response-oriented utilities. The new scattering recovery comparison fixes acquisition to isolate inference, and the real-video refinement does not test closed-loop acquisition. These implementations need not share the same design approximation.

Information boundary and evaluation. Inference and planning receive acquired records, public operators and protocols, and the supplied calibration and frozen assets declared for each comparison. Forward simulation of a candidate field is permitted; the object's unknown physical parameters and unacquired response values are not inputs. Evaluation uses saved physical properties without refitting to withheld responses. The reported response scores do not establish calibrated online uncertainty or stopping rules.

B Analytical Illustration: a Two-Segment Beam

For a unit simply supported Euler–Bernoulli beam, let the positive bending stiffness be d_1 on $[0, 1/2]$ and d_2 on $(1/2, 1]$. A point load at a produces the normalized moment $\bar{M}_a(s) = (1-a)s$ for $s \le a$ and $a(1-s)$ otherwise. With $G(r,s) = \min(r,s)[1 - \max(r,s)]$, the forward and observation operators are

$$F(\theta; a)(r) = \bar{w}(r) = \int_0^1 G(r,s)\frac{\bar{M}_a(s)}{d(s)}\,ds, \qquad H_{\eta,o}x = x'(o+\eta), \quad \theta = (d_1, d_2). \tag{18}$$

Here η is a persistent sensor-location error shared across measurements. Set $r = o + \eta$ and $A_j(a, r) = \int_{\Omega_j}(\mathbf{1}_{s>r} - s)\bar{M}_a(s)\,ds$ for the two segments. The ideal reading is

$$\mu_{a,o}(\theta, \eta) = A_1(a, o+\eta)/d_1 + A_2(a, o+\eta)/d_2. \tag{19}$$

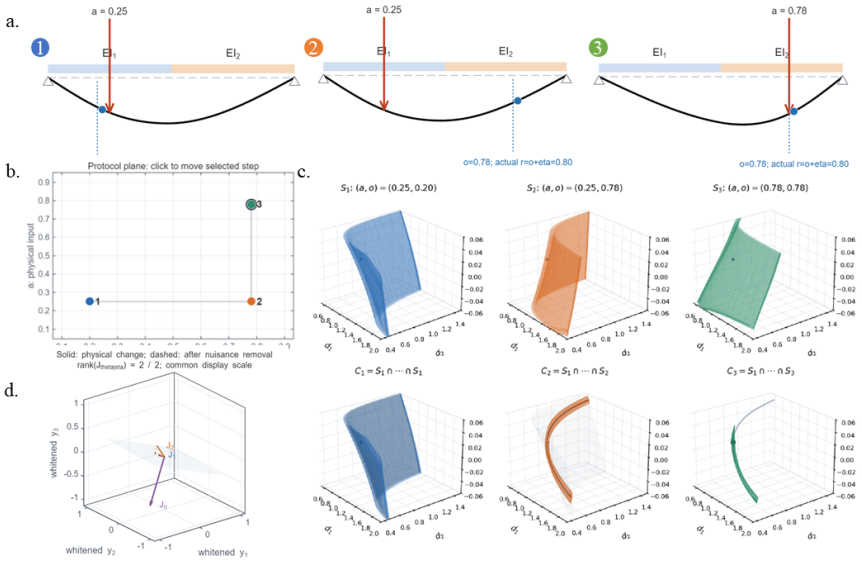


Figure 5: **Analytical geometry of a two-segment beam.** Changing the sensing location and then the load adds complementary constraints on the same stiffness and sensor-location parameters. The panels show the experiment sequence, individual compatibility sets, their intersections, and local sensitivity after nuisance projection. The EI_1, EI_2 labels correspond to d_1, d_2; $J_{\theta|\eta}$ in the local panel denotes J_{eff} in Equation (6).

At fixed η, one reading is a line in flexibility coordinates $(1/d_1, 1/d_2)$. Varying the unknown η sweeps a surface. Two independent readings generally determine the two flexibilities as functions of η, leaving a curve; a third complementary reading can cut this curve further.

The example uses $(d_1, d_2, \eta) = (1, 0.65, 0.02)$ and protocols $(a, o) = (0.25, 0.20), (0.25, 0.78), (0.78, 0.78)$. Differentiation gives $J_\theta = [-A_1/d_1^2, -A_2/d_2^2]$ and $J_\eta = -\bar{M}_a(r)/d(r)$ away from the material interface. Under unit noise, the stacked nuisance-projected sensitivity has smallest singular value approximately 5.21×10^{-3} at the nominal point. Complete noiseless deflection and curvature under a nonzero moment would instead permit $d(s) = -\bar{M}_a(s)/\bar{w}''(s)$; the present illustration concerns sparse, uncertain sensing.

C Thermal Reconstruction: Protocols and Physical Reuse

C.1 Physical variables, prior, and computational update

The generated heat-conduction problem reconstructs an interface-resistance field R and its depth d. The truth field has 32×32 cells; inference uses 16×16 coefficients and the frozen upsampling rule, giving 257 physical coordinates. Three emissivity coordinates represent the camera-side uncertainty shared by the full history. Heating controls the physical response F, and camera sampling controls H. The same unknown object and emissivity vector persist between acquisitions.

A frozen inverse network initializes a surrogate joint fit. Its highest-weight component gives the original estimate and initializes actual-F/H refinement. In the fixed transformed coordinates z for (θ, η), refinement approximately solves

$$\hat{z}_t \simeq \arg\min_z \left\{ \frac{1}{2} \sum_{(a_i, o_i, y_i) \in \mathcal{D}_t} \left\| R_{o_i}^{-1/2} \left[y_i - H_{\eta(z), o_i} F(\theta(z); a_i) \right] \right\|^2 - \log p_{0,z}(z) \right\}. \tag{20}$$

Here $p_{0,z}$ denotes the prior density in the declared inference coordinates z. A joint Gauss–Newton/Laplace approximation retains physical–nuisance cross-covariances. The planner evaluates the local utility in Appendix A.4, Equation (17) for three physical target curves, using predicted

response scales with a fixed 1 mK floor. Target sensitivity has no direct nuisance derivative, but its uncertainty depends on the joint covariance.

C.2 Matched acquisition and held-out objects

Every method shares prefix $(6597, 3037, 1539)$. Fixed adds protocol 3181; Universal fixed adds 3277; feedback chooses from 13 frozen joint heating–sensing candidates, including both controls. The Fixed sequence was designed on TRAIN from a 13,824-protocol space; Universal was selected on DEV. Each method uses four paid acquisitions, including the initial measurement, with total incident energy 0.16 J. The canonical initialization, complete-history endpoint, prior, and noise model are shared. Physical refinement is capped at twelve iterations. Equal acquisition cost does not imply equal computational cost.

The confirmation uses four new objects (seed 872027, indices 4096–4099), each with one camera/noise condition and disjoint from TRAIN/DEV generation seeds. Methods were frozen before transfer, and estimates were saved before evaluation. Feedback choices are $(3277, 3277, 1828, 1828)$; relative to Universal, target-error reductions are $(0, 0, -0.00032, +0.00856)$. This is one adaptive decision per object, $K = 3 \to 4$.

C.3 Recovery and physical-response metrics

Field error is $\mathrm{RMSE}(\hat{R}, R^\star)/R_{\max}$ with $R_{\max} = 1.2 \times 10^{-3}$ m^{2}K/W; depth error is $|\hat{d} - d^\star|$ in mm. The same saved properties predict interface-temperature-drop curves through F, without response fitting. Three target curves and three policy-unused tasks are evaluated at 141 times from 0.2 to 3 s. For each task, NRMSE divides curve RMSE by the larger of its true RMS and 1 mK, then averages tasks and objects equally.

The policy-unused tasks comprise two new heating loads and one internal ROI under an already observed load (action 137). They are excluded from the planner utility. Their DEV results were available during development; target definitions and TRAIN-object responses were used in task-model training. The held-out quantities are the confirmation objects' unmeasured responses. Cross-H additionally tests the frozen observation parameters at two withheld sensing protocols.

C.4 What the additional measurements distinguish

Two alternatives $\theta_\pm$ are fixed along a task-relevant direction at the common $K = 3$ anchor. For a history of K paid acquisitions, let M_K stack its predicted readings and $\mathcal{R}_K$ be its fixed observation covariance. We evaluate

$$d_K(\theta') = \min_{\eta \in \mathcal{E}} \left\| \mathcal{R}_K^{-1/2} \left[M_K(\theta', \eta) - M_K(\bar{\theta}, \bar{\eta}) \right] \right\|_2 . \tag{21}$$

The bounded nuisance vector is shared across the history. Alternatives, anchor, and bounds stay fixed. The fixed-reference diagnostic uses the shared $K = 3$ calibration estimate $\bar{\eta}$.

Mean separation gains $d_4 - d_3$ are 8.15, 24.22, and 24.02 for Fixed, Universal fixed, and feedback. Method-wise mean profiled-to-fixed-reference ratios range from 0.97 to 0.99, indicating limited nuisance compensation along this tested direction. For the exact profiled distance, retaining earlier records and their noise model makes d_K nondecreasing; numerical profiling approximates this minimum. The relevant evidence is separation magnitude at matched budget. Earlier-prefix values retrospectively evaluate alternatives defined at $K = 3$.

Table 4: Frozen thermal response errors (mK). Cross-H here is observation-space prediction error; it differs from scattering cross-H field variation.

Protocol	Target	New task	Cross-H
Fixed	0.08372	0.10323	0.14007
Universal fixed	0.06648	0.08936	0.17163
Feedback	0.06500	0.07799	0.08040

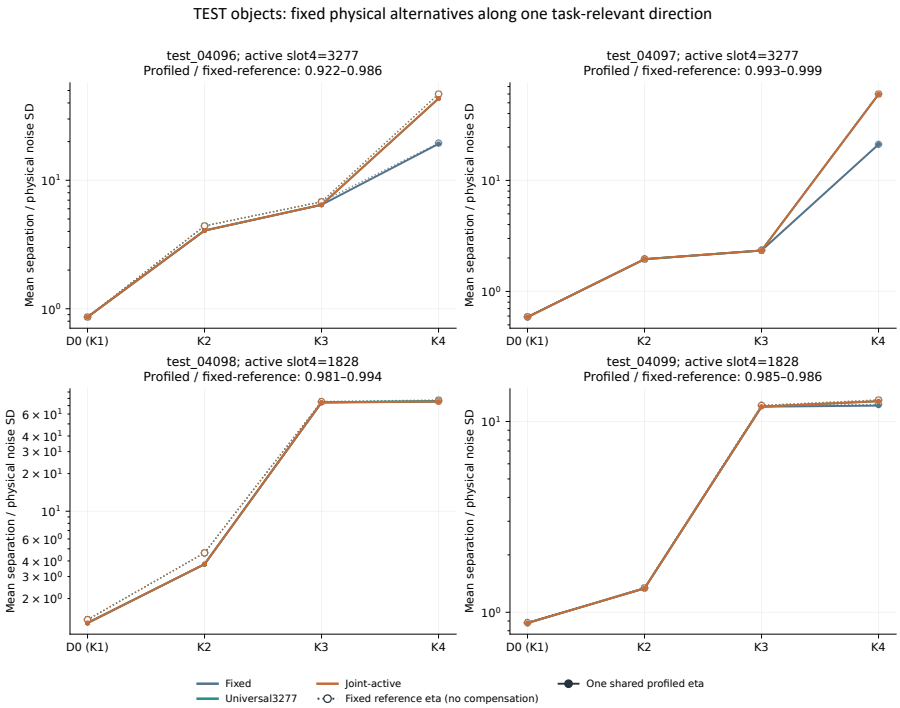


Figure 6: Complete thermal direction diagnostic, one TEST object per panel. Solid curves profile shared nuisance; dotted curves fix its estimated reference. Universal 3277 and Joint-active denote Universal fixed and Feedback, respectively. Distances are reported in physical-noise units.

Same-history refinement lowers new-task RMSE from 0.62199 to 0.10323 mK. At the common endpoint, feedback improves new-task prediction while field error rises: physical directions contribute differently to $\delta Q \simeq J_Q \delta\theta$, where $J_Q = \partial_\theta Q$ at the reference estimate.

D VIBRATION TOMOGRAPHY: EXTERNAL SOLVER AND EVIDENCE TRANSFER

D.1 PUBLIC-DATA PROTOCOL AND RECONSTRUCTION MODULES

The public damped-cube trajectories (Feng et al., 2022a;b) describe 50-mm cubes with fixed bottom boundaries and Poisson ratio 0.3. Each packet contains x/z displacements at 25 surface points: 6,000 samples over 3 seconds. Noise has standard deviation $10\,\mu$m and AR(1) coefficient 0.8; seed 1729 is specialized by object/action. The common prefix is back; two distinct additions come from left, right, and twist. Both systems pay for three complete packets. Public calibration, damping, geometry, and the mean-stiffness scale are shared.

VVT-adapted. The official alternating core is fixed at commit `4757f8298b29`. It estimates free stiffness/density fields on 8^3 elements. Adaptations comprise a motion-to-mode frontend, an equivalent sparse solve, and shifted NNLS with $E \geq 9$ Pa and $\rho \geq 1.27$ kg/m^3. At most ten modes pass realness ≤ 0.25; $\alpha_u = 1$ and stiffness/density smoothing is $10^{-10}/N$ and $10^{-7}/N$. Stopping uses 150–300 iterations, relative material changes below 10^{-3} for three consecutive iterations, or 3,000 seconds. The DEV-fixed sequence is back/left/right. This identifies the adapted public-motion baseline separately from the native raw-video pipeline.

Our material reconstruction and planning. Six inclusion bounds, two log contrasts, and one background-density coordinate define positive materials on 10^3 elements, with mean stiffness 9000 Pa. Fitting uses the first 1,000 samples of each fully charged packet, spectral initialization, a 10^{-4} regularizer, and L-BFGS-B with at most 64 objective calls. Planning uses a local Gauss–Newton covariance and scale-projected material sensitivity. A final high-contrast restart adds $\log 10$ to the estimated stiffness contrast and permits 64 further calls; the acquired-data objective selects the output. This classical fit uses a structured prior matched to the inclusion family.

D.2 METRICS AND THE UNCHANGED-CORE COMPARISON

On the common grid of $N = 10^3$ material cells, let $\ell = [\log(\widehat{E}/E), \log(\widehat{\rho}/\rho)]$ concatenate the material log-ratios and let $\bar{\ell} = (2N)^{-1} \sum_{j=1}^{2N} \ell_j$. Removing one joint scale gives

$$R = \left[\frac{1}{2N} \sum_{j=1}^{2N} (\ell_j - \bar{\ell})^2 \right]^{1/2}, \qquad P = \frac{\|u_{\mathrm{pred}} - u_{\mathrm{ref}}\|_2}{\|u_{\mathrm{ref}}\|_2}. \tag{22}$$

The stiffness and density components use this same scale; log-RMSE is not percentage error. Scale removal is evaluation-only. Response prediction uses the original frozen materials, candidate-material static preparation, and zero release velocity. It scores all 6,000 frames, 1,331 nodes, and three components. VVT propagates on 8^3 before Q1 displacement interpolation; material remeshing is used only for R. No amplitude or phase alignment is applied.

Replay retains these VVT reconstruction rules and changes only the purchased packet history. It started with object 04; a predeclared branch rule expanded to 07/10/05 after R or P improved by at least 1% on 04. This outcome-gated coverage is a case-level comparison. On 05/10, packet identities, modes, scientific iteration traces, and materials agree exactly; timing fields are excluded. On 07, response error improves even though material error increases slightly. The replay isolates evidence transfer through the fixed VVT reconstruction pipeline.

D.3 COMPLETE COMPARISONS AND ACQUISITION CONTROLS

DEV uses 01/03; the comparison uses 04/05/06/07/10/11. All have historical development exposure. The full-system comparison changes representation, mesh, fitting window, inference, and acquisition, with acquisition cost matched but computation unmatched. Our six nominal endpoint refinements exhaust their budgets; VVT reaches material stability on three objects and its iteration cap on three. Twist is an admissible acquisition, and its response is evaluated only for objects on which neither system purchases it. The response and replay comparisons use the same six physical objects.

Table 5: All six nominal-observation objects. VVT denotes VVT-adapted. Twist is scored only where neither system purchased it; dashes are inapplicable.

	Material R		Front P		Twist P	
Object	VVT	Ours	VVT	Ours	VVT	Ours
04	1.14223	0.22433	0.49066	0.05977	—	—
05	1.12401	0.24839	0.77232	0.00819	0.55550	0.00671
06	1.22318	0.44183	1.33563	0.13576	2.00947	0.09193
07	1.15358	0.15763	1.64641	0.02905	—	—
10	0.99396	0.55046	0.39539	0.10686	0.31630	0.07672
11	1.08524	0.57001	0.51050	0.12986	0.85218	0.10497

Table 6: Same-core VVT replay, including exact-history reuse controls. Reused rows are controls, not new experiments. Each row keeps the VVT configuration fixed.

		Material R		Front P	
Object	Replay / history	Fixed	Bridge	Fixed	Bridge
04	new / different set	1.14223	1.13410	0.49066	0.38468
07	new / different set	1.15358	1.15872	1.64641	0.34668
05	new / same order	1.12401	1.12401	0.77232	0.77232
10	new / reordered	0.99396	0.99396	0.39539	0.39539
06	reused / same order	1.22318	1.22318	1.33563	1.33563
11	reused / same order	1.08524	1.08524	0.51050	0.51050

D.4 Known-observation robustness

Known x/z gains $1.25/0.8$ preserve measurement-side noise but change corrected noise covariance. On object 04, our front error rises from 0.0598 to 0.3115 and material drift is 0.3658 (VVT: 0.0835); on 05, it rises from 0.0082 to 0.0119 and drift is 0.0545 (VVT: 0.8611). For 04's changed-gain readout, frozen nominal materials give error 0.06436 versus 0.19351 after refitting. Observation robustness is therefore not uniform.

D.5 Raw real-video extension: shared frontend, retained initialization

The public gelatin specimen contains a documented clay inclusion (Feng et al., 2022a, Sec. 6.2), but no registered internal material or defect labels are available here. Three original AVI files (8.80 GB; 34,809 processed frames) use shared official phase-motion and smoothing routines, with streamed fixed-bin extraction matching cached frequencies and modes to floating-point precision. Both methods retain clicks, camera calibration, $49 \times 47 \times 45$ mm geometry, and a 10^3 Q1 mesh. Official solver source is unchanged; verified Windows Q1 assembly replaces DOLFIN execution.

Front/right-pluck dynamics alone form the fitting packet. Left-side-twist dynamics are excluded from material fitting and checkpoint selection; static test-view calibration, official spectral bins, and mode identities remain known. Test modes 0/1 overlap fitting modes; test mode 2 (unique ID 3) is absent. Our classical refinement starts from VVT and optimizes 16 log-material coefficients with frequency weight 10, four subspace updates, and at most 150 L-BFGS-B iterations each. The lowest exact-FEM fitting objective selects the checkpoint, including a Native fallback (raw selection: stage 3). Local logs and matching recipe hashes place freezing before the first recorded held-out evaluation.

Frozen materials generate eigenmodes independently of test dynamics. Evaluation pairs observed and predicted modes among the first 20 predictions by Hungarian matching with cost $|\log(f_{\mathrm{pred}}/f_{\mathrm{obs}})| + 1 - \mathrm{MAC}$; matching does not update materials. Table 2C reports all three modes. Their individual shape errors are $(0.2806, 0.1997, 0.5450)$ versus $(0.3155, 0.1797, 0.3519)$; all three frequency errors worsen. Figure 3b shows that unseen mode; its numeric callouts average all three modes. Display alignment changes no metric. Native reconstruction costs 267.0 s and refinement adds 31.5 s (298.5 s total); shared three-video frontend time is 803.1 s, excluding download, packet preparation, and scoring.

Cached leave-one-mode-out shape errors improve in 6/6 modes (mean reduction 22.5%); mean frequency error falls from 6.539% to 6.047%, but worsens in 4/6 modes. The alternate front-held-out split worsens shape error; the right-held-out split improves both metrics. A clean-output replay of the raw entrypoint in the same installed environment reproduces the main metrics and recorded material/prediction hashes in 1292.6 s. The extraction, cached splits, and replay use the same specimen and software environment. This experiment demonstrates material refinement and frozen-property modal prediction through the official real-video frontend.

E WAVE-SCATTERING TOMOGRAPHY: FROZEN PRIORS, RECOVERY, AND COMPATIBILITY

E.1 PHYSICAL MODEL, DATA PERMISSIONS, AND COHORTS

The InverseBench linear operator maps $\theta \in [0,1]^{128\times128}$ to complex responses using its native finite-distance Hankel geometry, conjugation, and cell-area scaling (Zheng et al., 2025). Settings are $L_x = L_y = 0.18$, `wave`=6, radius 1.6, 20 transmitters, and 360 receivers. A history-shared receiver gain is

$$g_\eta(\phi) = \exp[\beta_c c(\phi) + \beta_s s(\phi) + i\{\gamma + \delta_c c(\phi) + \delta_s s(\phi)\}], \tag{23}$$

where $\eta = (\beta_c, \beta_s, \gamma, \delta_c, \delta_s)$, and c, s are receiver-centered cosine and sine. The measurement operator acts as

$$[H_{\eta,o}(x)]_r = g_\eta(\phi_r) x_r \tag{24}$$

for receiver indices r selected by o; with all receivers selected, we abbreviate it as H_η. Bounds are $\pm(0.3, 0.3, 0.4, 0.2, 0.2)$; conditions are zero, $(0.15, -0.08, 0.20, 0.12, -0.10)$, and its negative. The receiver-wise geometric mean of $|g_\eta|$ is one. Complex noise of RMS $\sigma_y = 10^{-4}$ is added after H, with the same noise bank across methods and conditions.

The new confirmation uses paid transmitters $\mathcal{T}_{\text{acq}} = \{0, 4, 10, 16\}$ and unacquired evaluation transmitters $\mathcal{T}_{\text{eval}} = \{3, 8, 13, 18\}$: 1,440 paid complex readings per history. Subscripts on F and y select and stack the indicated transmitter records.

Blind recovery receives paid records, public F/H models, configurations, and the shared frozen `inv-scatter-5m.pt` EMA checkpoint (EDMPrecond/DhariwalUNet). True fields, true H, and unacquired response values are evaluation-only. Network weights are not optimized; no new pretraining or policy learning is performed. Prior training-set membership and near-duplicate exclusion are unverified. The new roster is indices 68–91 ($n = 24$), three H conditions, one optimization/noise seed, and five methods (360/360 complete). Methods and output rules were frozen before those objects were opened. The known-H control reuses indices 68–75 ($n = 8$, 120/120), not a further independent cohort.

E.2 ACTUAL NEURAL AND PHYSICAL UPDATE PATHS

DPS uses 1,000 steps with guidance 560. R_0 appends alternating shared-H calibration and box-constrained field updates, at most 8×200 steps, with anchor scale $\tau = 0.1$. Its field subproblem has objective

$$\text{NLL} + \frac{\|\theta - c\|^2}{2\tau^2}, \qquad \text{NLL} = \frac{\|H_\eta F_{\mathcal{T}_{\text{acq}}}\theta - y_{\mathcal{T}_{\text{acq}}}\|^2}{\sigma_y^2}. \tag{25}$$

Starting from R_0, R_s applies five denoising/physics stages. R_+ replaces each denoiser call by fixed four-rotation averaging in normalized coordinates $u = 2\theta - 1$:

$$\overline{D}_\sigma(u) = \frac{1}{4}\sum_{k=0}^{3} T_k^{-1} D_\sigma(T_k u), \qquad c = \text{clip}_{[0,1]}\left[\frac{\overline{D}_\sigma(u) + 1}{2}\right]. \tag{26}$$

T_k rotates only the denoiser input, undoing the rotation before averaging; F, H, coordinates, and measurements do not rotate. The five σ values are $(0.8, 0.4, 0.2, 0.08, 0.03)$, with physical anchor scales $\tau = \sigma/2$. Each stage uses at most 3×100 field steps alternating with shared-H fitting. A final

original R_0-style physical repair anchors to the fifth-stage result. The unique last repaired endpoint is returned; no TEST-time selection among stages occurs. Changing centers is not a monotonicity guarantee for one global MAP objective. The complete R_+ uses 1,020 prior calls (1,000 inherited plus 20), not 20 standalone calls.

RED-diff retains the native algorithm core with a paid-complex-data/blind-H adaptation: 1,000 steps, learning rate 0.04, constant $\lambda = 0.0005$, and observation weight 1,500 (Mardani et al., 2024). It was selected from a bounded DEV shortlist, not an exhaustive optimum. Stronger DEV DiffPIR with $\lambda = 4$ and CG cap 240 reached the registered tolerance in 400/400 solves without outperforming the retained recovery; its weakness cannot be attributed solely to the old truncated-CG implementation. This is a controlled $K = 4$ extension, not the native 20-transmitter leaderboard.

E.3 Metrics, full costs, and paired confirmation

With $e(u, v) = \|u - v\|/\|v\|$,

$$E_\theta = e(\hat{\theta}, \theta^\star), \qquad E_H = e(g_{\hat{\eta}}, g_{\eta^\star}), \qquad Q_F = e\Big(F_{\mathcal{T}_{\text{eval}}}\hat{\theta}, F_{\mathcal{T}_{\text{eval}}}\theta^\star\Big). \tag{27}$$

For E_H, each g_η is evaluated at the common receiver locations. Q_{HF} additionally applies the respective frozen H. PSNR/SSIM use the fixed evaluator with range 1 and no truth alignment. C_H averages symmetric field distances across the three condition pairs; it is not a recovery or identifiability certificate. Means first average conditions within physical objects. Paired intervals use 10,000 object-bootstrap draws without multiplicity correction.

Table 7: Additional new-confirmation metrics and full deployment costs. Prior counts include parent inference; r_D is mean acquired-data residual in the stated noise normalization. All rows share fixed $K = 4$.

Recovery	$10^3 E_H \downarrow$	$Q_{HF} \downarrow$	r_D	C_H	Prior calls	Total s
DPS	5.9424	0.08263	9.09799	0.11390	1,000	35.12
R_0	0.6273	0.07392	0.68652	0.10779	1,000	41.67
R_s	0.8838	0.06700	1.22841	0.06157	1,005	48.13
RED-diff	0.8833	0.08902	1.92829	0.02837	1,000	21.79
R_+	0.5744	0.05954	0.65995	0.05226	1,020	54.04

$R_+ - R_0$ differences are $+1.6996$ dB PSNR (95% interval $[1.1740, 2.0586]$), $+0.02045$ SSIM ($[0.01433, 0.02641]$), -0.05052 field error ($[-0.05886, -0.04046]$), and -0.014394 Q_F ($[-0.016978, -0.011412]$). Each improves on 23/24 objects; object 91 is retained as the common failure case. E_H improves on 18/24. Relative to R_s, Q_F decreases by 11.11%; relative to RED-diff, PSNR increases by 3.0947 dB, but runtime is $2.4805\times$ and C_H is worse on 21/24 objects. Cost includes all parent inference; 12.37 s is only the increment over R_0.

Recovery under known observation parameters. All five methods receive the same true H on the eight-object subset. Within that control, R_+ improves the four recovery metrics over R_0 on 8/8 objects: $\Delta Q_F = -0.01647$ ($[-0.01990, -0.01328]$) and $\Delta\text{PSNR} = +1.9167$ dB ($[1.5745, 2.2572]$). Separately, all twenty same-method known-minus-blind mean-difference intervals (five methods, four recovery metrics) cross zero. This is neither a demonstrated average benefit nor evidence that knowledge is harmful or equivalent. $E_H = 0$ is imposed by the control, not learned calibration.

E.4 Compatibility: recovery selection versus new evidence

For an acquired history D containing M complex readings, define

$$r_D(\theta, \eta) = \frac{\|H_\eta F_D \theta - y_D\|^2}{M\sigma_y^2}. \tag{28}$$

Under independent circular complex Gaussian noise with $\mathbb{E}|\epsilon|^2 = \sigma_y^2$, the corresponding noise-resolution cutoff is

$$t_M = \frac{\chi^2_{2M,0.99}}{2M} \tag{29}$$

An attained legal fit below t_M establishes compatibility under this stagewise diagnostic.

Relaxing the receiver gain from the parametric family g_η to arbitrary complex gains shared across transmitters at each receiver gives

$$L_D(\theta) = \frac{1}{M\sigma_y^2} \sum_r \left(\|y_r\|^2 - \frac{|f_r^* y_r|^2}{\|f_r\|^2} \right) \le \min_{\eta \in \mathcal{E}} r_D(\theta, \eta). \tag{30}$$

Here f_r, y_r span transmitters; zero f_r contributes $\|y_r\|^2$. A numerically guarded $L_D > t_M$ excludes a fixed field for every legal H. Otherwise, failure of finite fitting remains unresolved. These are stagewise noise-resolution diagnostics, not selection-adjusted statistical tests; unlike the retained intersections in Appendix A.1, they do not by themselves establish nested compatibility sets.

For the new 72 paired object/condition endpoints, both R_0 and R_+ are compatible even at the same true H. Their mean symmetric field distance is 0.15919 and unacquired F-prediction distance is 0.04078; these are distances between estimates, not errors to truth. Only one R_s endpoint (object 71, nominal) supplies a near-threshold true-H-excluded/unknown-H-compatible witness. The main recovery gains therefore do not require observation compensation or new information.

Historical fixed-alternative evidence. The separate post-hoc DEV cohort (6–9) retains 84 fields: eight initial estimates, 64 local perturbations, and twelve primary alternatives. Table 8 reports all primary alternatives. Seven on three objects are compatible only after permitting unknown H relative to the true-H diagnostic; their minimum distance from the initial estimate is 0.160. All seven are excluded by later tested histories, including strong fixed $K = 4$ controls. Among the five non-witness primary alternatives, three are already compatible at true H and two remain excluded even after allowing legal H; all 64 local perturbations remain unresolved initially.

Table 8: Historical DEV alternatives, normalized by stage-specific cutoffs. Known: exact initial residual at true H; legal: attained upper bound with unknown H; new: relaxed lower bound after one acquisition. Stars mark seven compensation witnesses. S/E denote single/ensemble sources.

Alternative	Known	Legal	New	Alternative	Known	Legal	New
S6-9	0.998	0.723	31.07	E7-9*	8.186	0.629	25.74
S7-9	0.876	0.606	27.59	E7-10*	1.224	0.604	19.53
S8-9*	5.780	0.635	18.88	E8-9*	2.937	0.673	15.00
S9-9	0.594	0.582	8.17	E8-10*	5.865	0.662	15.08
E6-9	5.127	3.284	39.49	E9-9*	3.720	0.584	14.54
E6-10	9.962	3.567	39.74	E9-10*	1.261	0.594	12.32

Historical planning comparisons are not replaced by recovery gains. On indices 48–63 ($n = 16$), feedback with the original refined-DPS endpoint lowers field error from 0.2753 to 0.2730 relative to fixed $\mathcal{T}_{\mathrm{acq}}$, but worsens Q_F from 0.0792 to 0.0811. On indices 64–67 ($n = 4$), a two-seed, 400-step ensemble backend trades slightly better E_H (0.000772 to 0.000703) for worse Q_F (0.0705 to 0.0726) and C_H (0.0484 to 0.1279). The latter two worsen on all four objects; their paired differences are $+0.00217$ ($[0.00048, 0.00423]$) and $+0.07952$ ($[0.02295, 0.14034]$). Here C_H also includes action changes. Eighteen of 24 estimate–alternative pairs admit a common legal H; six are unresolved. These different backends/cohorts are not pooled with the new fixed-acquisition confirmation.