FROM PREDICTION TO MEASUREMENT: THE PROTOCOL BOUNDARY FOR MATERIAL PARAMETERS FROM VIDEO

Anonymous authors
Paper under double-blind review

ABSTRACT

Video-based inverse-physics systems increasingly report material parameters in physical units, yet predicting physical motion and measuring physical parameters are different tasks. For the continuum models these systems use, jointly scaling density and every stress-dimensioned parameter leaves the observed mechanics unchanged, so the observations determine gauge-invariant combinations such as the elastic speed scale $c = \sqrt{E/\rho}$ but not the absolute mass–stress scale. We turn this classical symmetry into a criterion for *when an observation protocol supports an absolute measurement*: the surviving gauge is exactly the subgroup that leaves invariant every exogenous datum the inference holds fixed. Kinematic and density-proportional forcing—passive motion, gravity, self-weight, and even displacement-controlled actuation—therefore preserve the gauge, while a fixed datum carrying an absolute mass or force scale removes it; with known geometry, a weighed mass suffices. The boundary governs evaluation as well. If every validation protocol preserves the gauge, every score computed from its observations is gauge-invariant, so held-out views, frames and kinematic interactions can validate predicted dynamics without validating the absolute parameters behind them: a candidate whose modulus is wrong by $100\times$, with density co-scaled, still attains silhouette IoU 1.0000, and one weighed mass exposes it. Across seven protocols, four constitutive families and released systems from two independent lineages, we find that when the mechanics leave the absolute scale undetermined it is supplied elsewhere—by an assumed density, a prior, an initialization or optimizer geometry—even where density is genuinely optimized. We state what to report, and how to validate, instead.

1 INTRODUCTION

Video-based inverse-physics methods increasingly recover not only motion but material parameters in physical units: recent systems couple 3D Gaussian Splatting (3DGS), neural radiance fields (NeRF) or related representations to differentiable simulators and report Young's modulus E, density ρ, yield stress, viscosity or a learned constitutive response (Xie et al., 2024; Zhang et al., 2024; Li et al., 2023; Cao et al., 2024; Chen et al., 2025). Accurate dynamics and accurate physical parameters are, however, different claims. We ask *when an observation protocol contains enough information for an absolute material parameter to count as a measurement*.

The observations may determine only an equivalence class. With geometry, boundary conditions and dimensionless constitutive parameters fixed, elastic motion depends on E and ρ only through the elastic speed scale $c = \sqrt{E/\rho}$, so $(E, \rho) \mapsto (\alpha E, \alpha\rho)$ leaves the ideal trajectory unchanged for every $\alpha > 0$. More generally, jointly scaling density and every stress-dimensioned parameter traces a *gauge orbit* of physically different materials that generate the same observation. More frames, more cameras, a better renderer or a larger network can pin c ever more precisely while leaving the orthogonal mass–stress scale exactly undetermined. Figure 1 shows this inside a released pipeline: making the material four times stiffer and four times denser leaves the rendered output unchanged to one 8-bit level, whereas changing density alone, which leaves the orbit, changes the motion at once.

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.

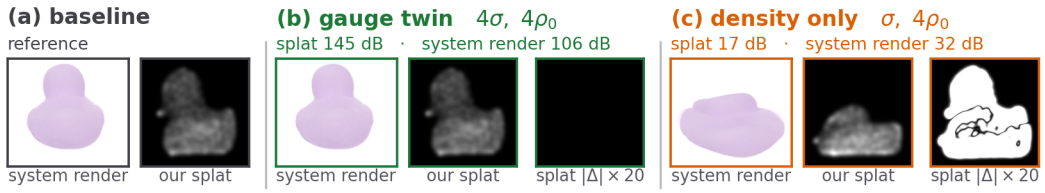

Figure 1: **Different absolute parameters, the same predicted video.** A released inverse-physics pipeline at the final frame under (a) its original parameters, (b) a *gauge-equivalent* setting with its learned stress response and ρ both $\times 4$, and (c) an off-orbit control with ρ alone $\times 4$. Each group shows the system's own render, an independent splat re-render, and the splat's difference from the baseline at $20\times$ gain. Setting (b) is a material four times stiffer yet agrees with the baseline to one 8-bit level in the system's renderer (PSNR 106 dB) and to 145 dB in the re-render; (c) does not (32 and 17 dB). The system is identified and audited in §5; supplementary video S1 shows all 151 frames.

A method can nonetheless return a unique absolute value, because something selects a point on the orbit. We distinguish three sources: a *physical anchor*, such as a weighed mass or a calibrated traction, which contributes mechanical information; a *statistical prior*, learned or assumed outside the observed mechanics; and a *selector*—a parameter bound, initialization, regularizer, parameterization or optimizer geometry—which picks a representative without evidence that it is the physically correct one. All three yield a number; only the first makes it a mechanical measurement. The question is therefore not whether a method outputs E, but what fixed its absolute scale.

The protocol determines the boundary. "Video cannot measure an absolute modulus" is too coarse: a recording may contain an independent mass or force reference. Mechanical identifiability depends on the forward model and the protocol, so we ask the converse of the usual fixed-experiment question: *given the absolute quantity we wish to report, which protocols can support it?* Our certificate (§4) answers it from the physical dimension of each datum the inference holds fixed. Passive motion, gravity and self-weight preserve the mass–stress gauge, and so, less obviously, does displacement-controlled actuation, since prescribing a motion does not prescribe the force that produces it. A calibrated force, pressure or mass removes the gauge; with known geometry, a weighed attached mass suffices. The scaling argument is classical (Bellman & Åström, 1970; Castro & de Boer, 2020); making the protocol the variable turns it from a diagnosis into a design rule.

Validation inherits the same boundary. If a protocol cannot distinguish two parameter settings, neither can any score computed only from its observations (§4.1). Held-out frames, additional views and even a new kinematic interaction can then validate predicted dynamics while carrying no evidence about the absolute scale: a modulus wrong by $100\times$, with density co-scaled, attains silhouette intersection-over-union (IoU) 1.0000 under a held-out kinematic interaction, and fails at once under a held-out interaction that carries a weighed mass. **Prediction success is not, by itself, parameter validation.**

Where the missing scale comes from. The certificate predicts that any absolute mass–stress quantity reported from gauge-preserving observations takes its missing coordinate from outside them, and released code bears this out (§5, §6). In one inherited lineage density is declared learnable but never stepped, and changing the assumed density moves the recovered modulus one-for-one at almost unchanged fit. In an independent lineage the gradient path is complete and density is genuinely optimized, yet its documented protocol contains no mass datum, so the scale remains conditioned on bounds, initialization and optimizer geometry. The mechanisms differ; the cause is the same. Our claim concerns the *mechanical* observation model, not RGB in general: appearance can support a statistical prior, which should then be reported as one.

Contributions. The identifiability machinery is established (Bellman & Åström, 1970; Castro & de Boer, 2020; Villaverde et al., 2019; Raue et al., 2009; Anguelova et al., 2012; Gross & Blüthgen, 2020); we use it to establish three consequences for video-based inverse physics.

- **A protocol boundary for absolute measurement.** A gauge certificate determines which mass-scale transformations survive a given continuum protocol, converting a non-identifiability observation into an experimental-design rule (§3–4).
- **A validity criterion for benchmarks.** Validation built entirely from gauge-preserving observations cannot establish a gauge-dependent parameter, however good the prediction; boundary-crossing interactions, not more gauge-preserving data, test the absolute scale (§4.1).
- **An account of how the missing scale is fixed in practice.** Across released systems from two lineages, four constitutive families and archival real video, we trace reported absolute scales to assumptions, priors, selectors or genuine physical anchors, and derive what to report and how to validate (§5–8).

2 Related Work

Video-based physical inference. PhysGaussian (Xie et al., 2024) simulates hand-specified materials on a static 3DGS with the material point method (MPM); inverse methods estimate parameters by analysis-by-synthesis (Li et al., 2023; Zhong et al., 2024; Cai et al., 2024; Cao et al., 2024), by distilling a video prior (Zhang et al., 2024), or feed-forward (Chen et al., 2025), on 3DGS (Kerbl et al., 2023) or NeRF (Mildenhall et al., 2020); dynamic and 4D reconstruction is likewise judged by agreement with held-out motion. Earlier, visual vibrometry (Davis et al., 2015) recovered the modulus-to-density ratio of rods from video and converted it to a modulus with measured densities, and differentiable video-to-physics identification predates 3DGS (Jatavallabhula et al., 2021). We ask not how to improve reconstruction or simulation, but which reported quantities a given observation model and protocol actually determine.

Identifiability and physical inverse problems. Scaling symmetries are a standard route to structural non-identifiability (Bellman & Åström, 1970; Castro & de Boer, 2020), although an absent symmetry does not establish identifiability (Villaverde & Massonis, 2021; Massonis & Villaverde, 2020). Related ambiguities are familiar elsewhere: inverse elasticity depends jointly on displacement, boundary conditions, density and forcing (Doyley, 2012); elastography images tissue stiffness (Ophir et al., 1991), and its shear-wave form converts a measured wave speed to a modulus through an assumed density (Sarvazyan et al., 1998); full-waveform inversion resolves velocity–density tradeoffs with extra physics or priors (Tarantola, 1984; Virieux & Operto, 2009); and identifiability and excitation design are formalized in systems biology and robot identification (Raue et al., 2009; Gautier & Khalil, 1990; 1992). We differ mainly in what is treated as the variable: rather than fixing an experiment and asking what it identifies, we characterize which *protocols* remove a specified gauge, which also decides which validation measurements can distinguish competing absolute values.

Recent work on identifiability from video. Recent work makes identifiability explicit: Wang et al. (2026b) study latent second-order dynamics from pixels, Rho et al. (2026b) address a metric-scale ambiguity, and Zhang et al. (2026) target full material parameters without presets; measurement-grounded benchmarks are also emerging (Khanbayov et al., 2026; Wang et al., 2026a). Our analysis is complementary: it isolates the mass–stress gauge of continuum models and asks, protocol by protocol, what datum removes it.

3 Gauge Structure of Visual Physical Inference

We first characterize what mechanical observations can distinguish, independently of any estimator. A body with Young's modulus E and density ρ, and fixed geometry, boundary conditions and dimensionless constitutive parameters, is observed through a calibrated renderer R as $V(t) = R(u(t))$, with u the displacement field. Small-strain elasticity gives $M\ddot{u} = -K(E)u$ with $M = \rho M_0$ and $K(E) = EK_0$, where M_0, K_0 depend only on those fixed quantities.

Proposition 1 (Passive mass–stress gauge)**.** *For any $\alpha > 0$, $g_\alpha : (E, \rho) \mapsto (\alpha E, \alpha\rho)$ leaves the passive trajectory invariant, $u(t; \alpha E, \alpha\rho) = u(t; E, \rho)$, and hence the rendered video. Passive mechanics constrain (E, ρ) only through $c = \sqrt{E/\rho}$; the orbit $\{(\alpha E, \alpha\rho)\}$ is unobservable.*

Modal frequencies depend on E/ρ and mode shapes on the fixed quantities alone (Appendix B). The point is not that E and ρ are hard to separate, but that every point on the orbit produces the same ideal observation.

Corollary 1 (Gravity does not remove the gauge)**.** *If gravity enters as a density-independent acceleration, $\ddot{u} = g + (E/\rho)f(u)$ is unchanged by g_α; written instead as a body force proportional to ρ, force and inertia co-scale. Known gravity does not identify the absolute mass–stress scale.*

Gravity can remove a different ambiguity—metric scale, given a calibrated timebase—but not this one (§4).

3.1 Beyond linear elasticity

Proposition 2 (Stress–density scaling)**.** *Let a continuum solver integrate $a = \rho^{-1}\nabla\cdot\sigma + g$ with g density-independent, let stress be homogeneous of degree one in the stress-dimensioned parameters θ_σ, and let every plastic return map use thresholds homogeneous of degree one in θ_σ. Then $(\theta_\sigma, \rho) \mapsto (\alpha\theta_\sigma, \alpha\rho)$ leaves the trajectory, and the rendered video, invariant for all $\alpha > 0$.*

Stress and density acquire the same factor, so the acceleration is unchanged and homogeneous thresholds preserve every return-map decision (Appendix B). Every parameter with nonzero *mass exponent*—the power of mass in its physical dimensions—therefore co-scales with density: Young's modulus, the Lamé constants, yield stress, cohesion, bulk modulus and dynamic viscosity survive only as E/ρ, σ_Y/E, c_{coh}/E, κ/ρ and μ_{visc}/ρ, while Poisson's ratio, friction angle, the fluid equation-of-state exponent and any relaxation time are untouched (Appendix Table 3).

A surviving parameter is not necessarily identifiable. Proposition 2 is a one-way test: a parameter moved by the gauge cannot be determined absolutely from gauge-preserving observations, but one it leaves untouched may still be unidentifiable or poorly conditioned—the friction angle is gauge-invariant, yet a sticky-floor shear drop moves the observable $319\times$ less than a granular column collapse. We therefore keep symmetry removal, distinguishability, identifiability and conditioning distinct throughout. The surviving combinations are also dimensional: expressing c in physical units needs metric scale and a calibrated timebase.

4 The Protocol Boundary

We now ask the experimental-design question: *which protocols remove such a transformation?* Let $G = \{g_\alpha\}_{\alpha>0}$ act on each quantity q as $q \mapsto \alpha^{m(q)}q$, where $m(q)$ is its mass exponent; thus density, stress, yield stress, dynamic viscosity, surface tension, traction, force and mass scale by α, while length, time, velocity, acceleration (including g), strain, Poisson's ratio and friction coefficients do not. For a protocol P, the *stabilizer* $G_P \leq G$ is the set of transformations leaving the complete forward problem unchanged, including the quantities the inference treats as known.

Proposition 3 (Protocol stabilizer)**.** *For a dimensionally homogeneous continuum model, G_P consists exactly of the g_α that leave every fixed, nonzero* exogenous *datum of P—a quantity supplied to the inference rather than generated by the motion—invariant. Hence (i) if P fixes no exogenous datum of nonzero mass exponent, $G_P = G$ and the gauge survives; (ii) a fixed nonzero exogenous datum of nonzero mass exponent reduces G_P to the identity.*

Kinematic control does not supply a force scale. Passive motion, gravity, self-weight, and prescribed displacement or velocity all preserve the gauge. The last case is easy to misread: under displacement control the actuator's reaction force is an *output* of the forward problem and co-scales with the material, so prescribing the trajectory anchors nothing—though measuring that reaction force in calibrated units would. Conversely an externally known force, pressure or added mass removes the gauge; with known geometry a weighed attached mass suffices, but if metric scale is also unknown a residual generator remains. The classification runs by dimension, not by name: ideal rigid contact and Coulomb friction preserve the gauge, whereas penalty contact with a fixed stiffness, or drag against a known ambient density, breaks it.

Gauge removal is necessary, not sufficient. Proposition 3 is algebraic. A datum that removes the symmetry must also couple to the realized motion for parameter values to be distinguishable;

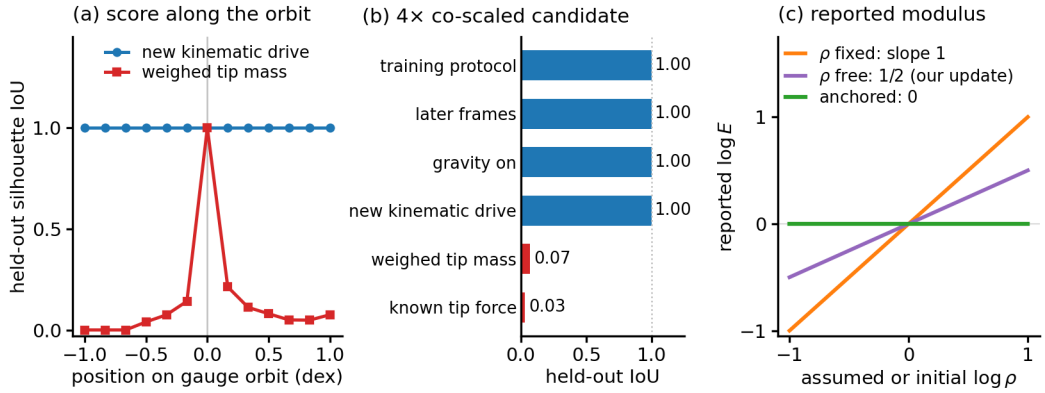

Figure 2: **Validation must cross the same boundary as estimation.** (a) Held-out silhouette IoU along the gauge orbit, i.e. for candidates with E and ρ co-scaled. Under a new kinematic drive a $100\times$ error in E still scores 1.0000; a held-out interaction carrying a weighed tip mass singles out the truth. (b) Across the protocol battery, only protocols carrying an independent mass or force scale detect a four-fold co-scaling (Appendix Table 5). (c) The same unresolved coordinate in estimation: with density fixed the fitted modulus follows the assumed density (measured slopes 0.976–1.003); with both free and no anchor the selected point depends on the optimizer (slope $1/2$ for our symmetric update, measured 0.489); an anchor removes the dependence (§5–6).

distinguishability does not imply identifiability, nor identifiability good conditioning; and anchor strength is continuous where symmetry removal is binary. In our design model, entrained air (Sader, 1998) shifts a steel rule's first frequency by -0.17%, so the symmetry is exactly gone, yet it bounds density only to ± 0.5 dex (one dex is a factor of ten), where one weighed mass reaches 0.004 dex; here and below $\log$ denotes $\log_{10}$. We use the proposition only to decide whether the gauge survives, and test distinguishability and conditioning separately.

Numerical verification. We simulate two worlds related by g_4 under seven protocols in our moving-least-squares MPM (MLS-MPM) solver: the four gauge-preserving protocols leave them identical to round-off ($\leq 4 \times 10^{-12}$), and the three carrying a calibrated force, pressure or added mass separate them (Appendix Fig. 4, Table 4); magnitudes on the broken side depend on the protocol, and the prediction is the classification.

4.1 Prediction success is not parameter validation

The same boundary governs evaluation.

Corollary 2 (Gauge-blind validation)**.** *Let a validation suite use protocols $P_1, \ldots, P_k$ with combined stabilizer $G_{\mathrm{val}} = \bigcap_i G_{P_i}$. For any $g \in G_{\mathrm{val}}$, every score $S = \Phi(F_{P_1}(\theta), \ldots, F_{P_k}(\theta), F_{\mathrm{obs}})$ satisfies $S(g\theta) = S(\theta)$, so the suite cannot establish any reported quantity that changes under g.*

The corollary is elementary; its consequence for benchmark design is not. Held-out frames, additional views, trajectory error and a second kinematic interaction can test prediction strongly while remaining exactly insensitive to the absolute mass–stress coordinate; to validate that coordinate, the validation protocol itself must cross the boundary. Figure 2 shows this directly: co-scaling E and ρ by $100\times$ leaves a new kinematic interaction at silhouette IoU 1.0000, while a held-out interaction that carries a weighed mass exposes the same error. An off-orbit control fails everywhere, so the metrics are sensitive—only not to the gauge direction (Appendix Table 5).

What existing validation can establish. Table 1 asks one narrow question of the evidence offered for each absolute output: does it contain a gauge-breaking datum? None of the evaluations offered for an absolute modulus does; the implementations are audited in §5. This does not make an absolute prediction from RGB meaningless—appearance can carry a material prior—but a gauge-preserving

Table 1: **Information supplied by the validation protocol.** For each system, the evidence offered for its material outputs and whether it contains an independent mass- or force-scale datum; "supported mechanically" is what those observations can constrain, given metric scale and timebase. PAC-NeRF's synthetic truth is generated at the density its fit assumes, so agreement on E is supplied by construction. Visual vibration tomography (VVT) is a positive control that reports relative fields under a declared prior. Full census and audit modes: Appendix Table 7.

System	reported quantity	validation evidence	gauge-breaking datum?	supported mechanically
PAC-NeRF (Li et al., 2023)	E, σ_Y, κ, μ	held-out rendering; parameter error at the assumed ρ	no	invariant combinations
NeuMA (Cao et al., 2024)	stress response at fixed ρ	render metrics, particle Chamfer distance	no	stress$/\rho$
Vid2Sim (Chen et al., 2025)	E	PSNR, SSIM, FoVVDP of reconstruction and future-state prediction	no	dynamics plus learned prior
M-PhyGs (Wada et al., 2026)	E, ρ per segment	held-out and cross-sequence PSNR, IoU, Chamfer	no	E/ρ per segment
VVT (Feng et al., 2022) *(control)*	relative fields	eigen-residual, declared scale prior	declared prior	relative fields

mechanical validation cannot establish the missing coordinate, and a reported scale must then come from another source, which should be named.

Multiple gauges can coexist. With the scene size L also unknown, the passive problem has two null directions in $(\log E, \log \rho, \log L)$, $u_1 = (1, 1, 0)$ and $u_2 = (2, 0, 1)$. A known g removes u_2, given a calibrated timebase, but leaves u_1 at machine zero (8.4×10^{-28}), so gravity anchors metric scale without anchoring a modulus (Bieler et al., 2019); protocol design should be stated generator by generator (Appendix B.1).

5 Where the Absolute Scale Comes From in Released Systems

The certificate predicts that a system reporting a gauge-dependent absolute quantity from gauge-preserving observations must fix the missing coordinate outside them. We test this on released code, checkpoints, assets and solvers. The audits serve different purposes—testing estimators, testing the gauge inside a released forward model, and identifying the source that selects the reported scale—and Appendix Table 7 gives the audit mode for each.

Positive control: a declared gauge choice. VVT (Feng et al., 2022) states the modulus–density scaling ambiguity, regularizes the scale toward a declared prior, and reports relative material variation. Fixing the gauge with a prior is not itself a problem; what matters is whether the source is declared and the quantity interpreted accordingly.

Fixed density selects the absolute modulus. PAC-NeRF (Li et al., 2023) optimizes density in none of its 25 shipped configurations, all of which use $\rho = 1000$. When we re-run its released fit and change only that constant over a $16\times$ range, the recovered modulus moves with log–log slope 1.003 ± 0.006 against the gauge prediction of 1, while the attained fit changes by under 5%: the observations fix E/ρ and the assumed density supplies the scale (Appendix Fig. 17). ProJo4D (Rho et al., 2026a) and MonoPhysics (Rho et al., 2026b) inherit the convention ($26/26$ and $4/4$ configurations), and re-running ProJo4D over an $8\times$ range gives slope 0.990 ($R^2 = 1.00000$) at 3.7% fit variation. These systems share ancestry, so we read this as one convention propagating, not as independent evidence of prevalence.

The same gauge inside a released prediction pipeline. The pipeline in Figure 1 is NeuMA (Cao et al., 2024). Applying g_4 to its released adaptor leaves the simulated trajectory unchanged to float32 precision (RMS 3×10^{-8}); through NeuMA's own rasterizer the output agrees with the baseline to one 8-bit level (106 dB), and to 145 dB in an independent splat re-render. The PSNR is finite, not infinite, because the solver is float32 and the frames are not bit-identical (Appendix A); density alone leaves the orbit and changes the prediction strongly. Feed-forward prediction does not escape the requirement either: Vid2Sim (Chen et al., 2025) predicts an absolute E at a fixed reference density

Table 2: **Three mechanisms that keep a learnable density fixed in one inherited lineage.** Each system declares density a learnable `nn.Parameter`, ships configurations at a fixed density, and has no density branch in its optimizer.

System	density gradient	observed at runtime
GIC (Cai et al., 2024)	complete	gradient populated on $194/195$ iterations
ProJo4D (Rho et al., 2026a)	none (not differentiable)	gradient `None` on $80/80$ iterations
MonoPhysics (Rho et al., 2026b)	cut by `.detach()`	—

and returns the same value for gauge-equivalent inputs whose true moduli differ $8\times$, so its absolute output reflects its training distribution and assumed scales (Appendix D).

Optimizing density does not make the scale observable. M-PhyGs (Wada et al., 2026), from an independent lineage, is the most informative case because density is genuinely optimized: its released solver returns a density gradient and the optimizer steps it. Yet its documented mechanical protocol is gauge-preserving—the human interaction overwrites particle velocities (a kinematic datum), gravity and the estimated initial internal force enter in proportion to particle mass, and neither the metric calibration nor the ground contact supplies a mass or force scale. The parameterization mirrors this, `particle_E = particle_normalized_stiffness` $\times$ `particle_density`: perturbing the density head by 0.53 dex moves ρ while E/ρ changes by 1.4×10^{-7}, at float32 precision, and its material regularizers are gauge-invariant. *Under the documented inputs we find no datum that fixes the global mass–stress scale, so the selected absolute density is conditioned on bounds, initialization, parameterization and optimizer path unless a further information source is identified.* Its validation is held-out and cross-sequence PSNR, IoU and Chamfer—evidence Corollary 2 shows to be blind to the orbit. Because its kernels pin an NVIDIA Warp version whose successor changed an API they rely on, for this system we report symbolic analysis, source inspection and parameter perturbation rather than an orbit rollout.

Summary. A fixed density converts an observed E/ρ into an absolute E; a learned predictor supplies scale from its training prior; an optimizer selects a point along an unobserved direction even with both variables free. None creates mechanical information absent from the protocol.

6 Freeing Density Does Not Identify the Scale

Making density trainable is not making it observable: if the mechanical loss is invariant along the gauge, freeing both E and ρ changes the optimization problem but adds no information about the missing scale.

The data gradient is orthogonal to the gauge. The observation depends on (E, ρ) only through $\log c^2 = \log E - \log \rho$, so the data loss is a function of that difference, its gradient in $(\log E, \log \rho)$ is proportional to $(1, -1)$, and its component along the gauge direction vanishes identically (measured: 4.6×10^{-15} of the gradient norm). An optimizer can still move along the gauge—through unequal learning rates, preconditioning, momentum, regularization, parameterization or initialization—but that movement is not determined by the data.

Fixing and freeing select the scale differently. With density fixed, fitting c gives $\log E = \log \rho_{\text{fixed}} + \log c^2$, so the modulus tracks the assumed density one-for-one (measured slope 0.976). With both free under our symmetric log-parameterized update, the gauge coordinate $s = \log E + \log \rho$ is conserved to 2.8×10^{-13} dex, so $\log E = (s + \log c^2)/2$ and the modulus has slope $1/2$ against the *initial* density (measured 0.489)—a property of that update rule, not a law; another optimizer selects another point on the same orbit. Freeing density changes the *source* of the absolute scale, from a fixed value to initialization and optimizer geometry; it does not make it data-determined.

6.1 What happens in released implementations

A gradient can exist without an update. On PAC-NeRF's `elastic/0` scene with GIC's shipped configuration, `log_rho.grad` is populated on 194 of 195 iterations, with magnitude up to 2.5×10^1, while $\log_{10} \rho$ stays at 2.698970 and $\log_{10} E$ optimizes normally: the gradient is computed, but density is absent from the optimizer step. Supplying the missing branch (density initialized at 1000) makes density move as soon as material optimization begins; over 40 material iterations $\log_{10} \rho$ rises $3.000 \to 3.504$ while $\log_{10} E$ falls $5.500 \to 4.992$, so the observable coordinate moves -1.013 dex and the gauge coordinate only -0.0037 dex—the optimizer adjusts what the observations constrain and leaves the absolute scale near its initialization. ProJo4D and MonoPhysics reach the same outcome through other mechanisms (Table 2), and M-PhyGs (§5) shows that the non-identifiability survives a complete gradient path.

Implication. Whether density is frozen, detached, omitted from the optimizer or actively updated is an optimization property; whether its value is determined by the observations is an identifiability property. Freeing an unobservable coordinate can change the reported number, but only a gauge-breaking information source makes that coordinate a measurement.

7 Verification in Simulation and Real Video

The classification holds across constitutive models. In our MLS-MPM solver the transformation of Proposition 2 leaves trajectories bitwise identical in these runs across four constitutive blocks, including von Mises and Drucker–Prager plasticity, while density-only, modulus-only and off-orbit perturbations move the observable orders of magnitude further; dimensionless parameters are untouched, as predicted (Appendix Table 3, Fig. 14). The ambiguity also survives real camera geometry: pushed through the calibrated cameras and masks of a released multi-view capture (Zhong et al., 2024)—a simulation scored against real observations, not a physical experiment—two gauge-equivalent materials differ by exactly zero silhouette IoU at every frame on all five objects, and the marginal for $\log E$ fills the declared 2-dex box while $\log c$ stays within one grid cell (Appendix C).

The same orbit in real high-speed video. Davis et al. (2017) released 2500 fps video of $1/8''$ metal rods with measured per-specimen densities. For a clamped rod of fixed geometry the fundamental frequency determines c, not E and ρ separately. Aluminium and steel form an approximate matched-speed pair: they differ $2.88\times$ in measured density and $2.90\times$ in reference modulus, so their speed scales agree to 0.3%, and their fundamentals fall in the same FFT bin while brass and copper resolve separately (Fig. 3a). A lock-in mode shape rising from clamp to tip confirms that the peak is the rod bending, not the camera moving (Fig. 3b).

An anchor converts the invariant into an absolute estimate. Combining each rod's video-derived c with its measured density through $E = \rho c^2$ gives moduli that track the references with log–log slope 0.971 ($R^2 = 0.987$; $n = 4$, descriptive) and median absolute error 0.027 dex (6.3%; Fig. 3c): the video quantity that cannot determine E becomes an absolute estimate once measured density supplies the missing coordinate. It also shows the limit. Propagating the density, diameter and length uncertainties covers the reference for three rods but misses steel by -4.6σ (Fig. 3d); steel's inputs are the most precisely known, so its interval is narrowest and exposes residual model error rather than hiding it. An anchor removes the structural gauge; it does not guarantee an accurate forward model (Appendix E).

Limitations. The classification assumes stress and plastic thresholds homogeneous of degree one in the stress-dimensioned parameters; learned and some rate-dependent laws need separate analysis. The rod references are manufacturer values for the alloys, not measurements of these specimens, so the experiment validates the inference procedure rather than providing metrological validation. Appendix F specifies a physical anchor-switch experiment that would close this gap; it has not been executed.

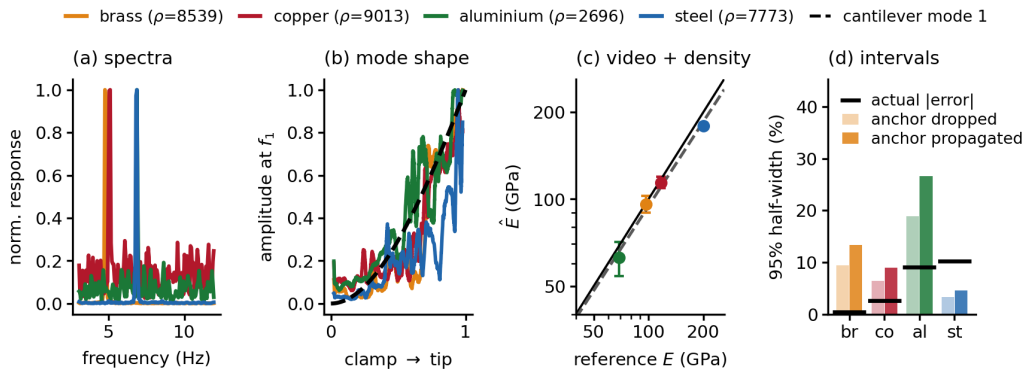

Figure 3: **The gauge orbit and a density anchor in real high-speed video** (Davis et al., 2017). (a) Aluminium and steel differ $2.88\times$ in density but share an elastic speed scale, and fall in the same FFT bin; brass and copper resolve. (b) The lock-in amplitude rises from clamp to tip along the first bending mode: the motion is the rod's, not the camera's. (c) Video-derived c plus each rod's measured density gives an absolute modulus (solid: 1:1; dashed: log–log fit, slope 0.971). (d) Nominal 95% half-widths with the anchor's uncertainty dropped and propagated, against the actual error: steel, whose inputs are known best, is the only rod whose error escapes its interval.

8 Implications for Reporting and Benchmark Design

The distinction drawn here is not between methods that do and do not report absolute parameters, but between the sources that make them absolute: a fixed density, a learned prior and a weighed mass can all select an absolute modulus, and they support different interpretations of it.

Report the observable and the source of the absolute scale. For each dimensional material parameter, report the combination the observations constrain, the additional quantity needed to express it absolutely, and the source of that quantity. Video constrains c under the forward model's assumptions; reporting $E = \rho c^2$ is then appropriate when ρ is measured or otherwise supplied, provided its source is stated and its uncertainty propagated: for a density independent of $\hat{c}$, $\mathrm{Var}(\log E) = \mathrm{Var}(\log \rho) + 4\,\mathrm{Var}(\log c)$, and treating an assumed density as exact silently drops the first term. A nominal 95% interval computed with $\rho = 1000$ then covers 24% of cases for water and 0% for granular media and foams, against 94% with a class density prior carried, and the damage grows as the video improves (Appendix Fig. 13). A statistical prior is not a physical measurement, but neither is it invalid; stating whether the scale came from the mechanics, a prior or a selector is what makes the claim interpretable.

Validate with a protocol that can distinguish the quantity. More frames, views or gauge-preserving interactions sharpen the observable dynamics without adding information along the missing direction. An absolute parameter should be validated with at least one protocol whose stabilizer excludes the corresponding gauge—for the mass–stress gauge, a calibrated force or pressure, or a known attached mass with known geometry. Crossing the algebraic boundary is only the first requirement: the anchor must couple to the motion, the inverse problem must be well conditioned, and the forward model must be adequate, as the steel rod shows.

Conclusion. Video-based inverse physics can determine physically meaningful material information without determining absolute material parameters. The object of inference is first the equivalence class the protocol supports; an absolute measurement requires information that removes the relevant gauge, and validating it requires a protocol sensitive to the same direction. We characterized this boundary for the mass–stress gauge across the constitutive models current systems use, showed that prediction can be correct while an absolute parameter remains unvalidated, traced where released systems obtain the missing scale, and showed with real video how a declared anchor converts the observable invariant into an absolute estimate. Identify what the observations constrain, state what supplies the rest, and test an absolute claim with an experiment that can distinguish it.

AI USE STATEMENT

In this work, we used generative AI tools to help develop and refine the research idea, the theoretical results and their proofs, to design and implement the experiments and code, to interpret results, and to draft and edit the manuscript, figures and references. We have not used generative AI tools to generate data or for translation, and qualitative data analysis is not applicable to this work. All AI-assisted work was reviewed and checked by one of the authors. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Milena Anguelova, Johan Karlsson, and Mats Jirstrand. Minimal output sets for identifiability. *Mathematical Biosciences*, 239(1):139–153, 2012. doi: 10.1016/j.mbs.2012.04.005.

ASTM International. Standard test method for dynamic Young's modulus, shear modulus, and Poisson's ratio by impulse excitation of vibration. ASTM E1876-22, 2022.

R. Bellman and K. J. Åström. On structural identifiability. *Mathematical Biosciences*, 7:329–339, 1970. doi: 10.1016/0025-5564(70)90132-X.

Didier Bieler, Semih Günel, Pascal Fua, and Helge Rhodin. Gravity as a reference for estimating a person's height from video. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019.

E. Buckingham. On physically similar systems; illustrations of the use of dimensional equations. *Physical Review*, 4(4):345–376, 1914.

Junhao Cai, Yuji Yang, Weihao Yuan, Yisheng He, Zilong Dong, Liefeng Bo, Hui Cheng, and Qifeng Chen. GIC: Gaussian-informed continuum for physical property identification and simulation. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

Junyi Cao, Shanyan Guan, Yanhao Ge, Wei Li, Xiaokang Yang, and Chao Ma. NeuMA: Neural material adaptor for visual grounding of intrinsic dynamics. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

Mario Castro and Rob J. de Boer. Testing structural identifiability by a simple scaling method. *PLOS Computational Biology*, 16(11):e1008248, 2020. doi: 10.1371/journal.pcbi.1008248.

Chuhao Chen, Zhiyang Dou, Chen Wang, Yiming Huang, Anjun Chen, Qiao Feng, Jiatao Gu, and Lingjie Liu. Vid2Sim: Generalizable, video-based reconstruction of appearance, geometry and physics for mesh-free simulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 26545–26555, 2025.

Abe Davis, Katherine L. Bouman, Justin G. Chen, Michael Rubinstein, Frédo Durand, and William T. Freeman. Visual vibrometry: Estimating material properties from small motions in video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5335–5343, 2015.

Abe Davis, Katherine L. Bouman, Justin G. Chen, Michael Rubinstein, Oral Büyüköztürk, Frédo Durand, and William T. Freeman. Visual vibrometry: Estimating material properties from small motions in video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4):732–745, 2017. doi: 10.1109/TPAMI.2016.2622271.

M. M. Doyley. Model-based elastography: A survey of approaches to the inverse elasticity problem. *Physics in Medicine and Biology*, 57(3):R35–R73, 2012. doi: 10.1088/0031-9155/57/3/R35.

Pierpaolo Dragonetti, Marco Civera, Gaetano Miraglia, and Rosario Ceravolo. A benchmark dataset for the validation of phase-based motion magnification-based experimental modal analysis. *Data*, 10(4):45, 2025. doi: 10.3390/data10040045.

Berthy T. Feng, Alexander C. Ogren, Chiara Daraio, and Katherine L. Bouman. Visual vibration tomography: Estimating interior material properties from monocular video. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

M. Gautier and W. Khalil. Direct calculation of minimum set of inertial parameters of serial robots. *IEEE Transactions on Robotics and Automation*, 6(3):368–373, 1990.

M. Gautier and W. Khalil. Exciting trajectories for the identification of base inertial parameters of robots. *The International Journal of Robotics Research*, 11(4):362–375, 1992.

Torsten Gross and Nils Blüthgen. Identifiability and experimental design in perturbation studies. *Bioinformatics*, 36(Suppl 1):i482–i489, 2020. doi: 10.1093/bioinformatics/btaa404.

Krishna Murthy Jatavallabhula, Miles Macklin, Florian Golemo, Vikram Voleti, Linda Petrini, Martin Weiss, Breandan Considine, Jérôme Parent-Lévêsque, Kevin Xie, Kenny Erleben, Liam Paull, Florian Shkurti, Derek Nowrouzezahrai, and Sanja Fidler. gradSim: Differentiable simulation for system identification and visuomotor control. In *International Conference on Learning Representations (ICLR)*, 2021.

Pierre Jop, Yoël Forterre, and Olivier Pouliquen. A constitutive law for dense granular flows. *Nature*, 441(7094):727–730, 2006.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3D Gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4), 2023. doi: 10.1145/3592433.

Rasul Khanbayov, Mohamed Rayan Barhdadi, Erchin Serpedin, and Hasan Kurban. IRIS: A real-world benchmark for inverse recovery and identification of physical dynamic systems from monocular video. *arXiv preprint arXiv:2603.16432*, 2026.

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 (ICLR)*, 2023.

Pingchuan Ma, Peter Yichen Chen, Bolei Deng, Joshua B. Tenenbaum, Tao Du, Chuang Gan, and Wojciech Matusik. Learning neural constitutive laws from motion observations for generalizable PDE dynamics. In *International Conference on Machine Learning (ICML)*, 2023.

Mohamed A. Mahmoud. Dynamic spring constants of higher resonance modes of uniform and tapered microcantilevers with tip masses. *Micron*, 129:102795, 2020. doi: 10.1016/j.micron.2019.102795.

Gemma Massonis and Alejandro F. Villaverde. Finding and breaking Lie symmetries: Implications for structural identifiability and observability in biological modelling. *Symmetry*, 12(3):469, 2020.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. In *European Conference on Computer Vision (ECCV)*, 2020.

Benjamin Muntz. Dimensional analysis is a gauge theory. *arXiv preprint arXiv:2607.21695*, 2026.

Patrick O'Donoughue, François Gautier, Erwan Meteyer, Thomas Durand-Texte, Mathieu Secail-Geraud, Félix Foucart, Olivier Robin, Alain Berry, Manuel Melon, Charles Pézerat, Adrien Pelat, and Pascal Picart. Comparison of three full-field optical measurement techniques applied to vibration analysis. *Scientific Reports*, 13:3261, 2023. doi: 10.1038/s41598-023-30053-9.

J. Ophir, I. Céspedes, H. Ponnekanti, Y. Yazdi, and X. Li. Elastography: A quantitative method for imaging the elasticity of biological tissues. *Ultrasonic Imaging*, 13(2):111–134, 1991.

Andreas Raue, Clemens Kreutz, Thomas Maiwald, Julie Bachmann, Marcel Schilling, Ursula Klingmüller, and Jens Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–1929, 2009.

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 (TMLR)*, 2026a. arXiv:2506.05317.

Daniel Rho, Jun Myeong Choi, Matthew Thornton, Biswadip Dey, and Roni Sengupta. MonoPhysics: Estimating geometry, appearance, and physical parameters from monocular videos. *arXiv preprint arXiv:2605.30320*, 2026b.

John E. Sader. Frequency response of cantilever beams immersed in viscous fluids with applications to the atomic force microscope. *Journal of Applied Physics*, 84(1):64–76, 1998. doi: 10.1063/1.368002.

Armen P. Sarvazyan, Oleg V. Rudenko, Scott D. Swanson, J. Brian Fowlkes, and Stanislav Y. Emelianov. Shear wave elasticity imaging: A new ultrasonic technology of medical diagnostics. *Ultrasound in Medicine and Biology*, 24(9):1419–1435, 1998.

Albert Tarantola. Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49 (8):1259–1266, 1984.

Alejandro F. Villaverde and Gemma Massonis. On testing structural identifiability by a simple scaling method: Relying on scaling symmetries can be misleading. *PLOS Computational Biology*, 17(10): e1009032, 2021. doi: 10.1371/journal.pcbi.1009032.

Alejandro F. Villaverde, Nikolaos Tsiantis, and Julio R. Banga. Full observability and estimation of unknown inputs, states and parameters of nonlinear biological models. *Journal of the Royal Society Interface*, 16(156):20190043, 2019. doi: 10.1098/rsif.2019.0043.

J. Virieux and S. Operto. An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 74(6):WCC1–WCC26, 2009.

Norika Wada, Kohei Yamashita, Ryo Kawahara, and Ko Nishino. M-PhyGs: Multi-material object dynamics from video. In *Findings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR Findings)*, 2026. arXiv:2512.16885.

Shuai Wang, Yaxin Feng, Xuekun Jiang, Shihan Tian, Ningyu Yan, Xing Shen, Chaoyang Lyu, Hui Wang, Yunsong Zhou, Hanqing Wang, Jiangmiao Pang, Yang Xiang, Xing Gao, Chunhua Shen, and Weinan Zhang. GAUGE: A measurement-grounded benchmark for physical fidelity in simulation engines and video world models. *arXiv preprint arXiv:2608.05948*, 2026a.

Yuanyuan Wang, Wenjie Wang, Kun Zhang, and Mingming Gong. Physics from video: Identifiability of time-invariant second-order ODEs under minimal trajectory conditions. In *International Conference on Machine Learning (ICML)*, 2026b. arXiv:2606.00115.

Tianyi Xie, Zeshun Zong, Yuxing Qiu, Xuan Li, Yutao Feng, Yin Yang, and Chenfanfu Jiang. PhysGaussian: Physics-integrated 3D Gaussians for generative dynamics. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

Klemen Zaletelj, Janko Slavič, and Miha Boltežar. Full-field DIC-based model updating for localized parameter identification. *Mechanical Systems and Signal Processing*, 164:108287, 2022. doi: 10.1016/j.ymssp.2021.108287.

Changshe Zhang, Jie Feng, Siyu Chen, Guanbin Li, Ronghua Shang, and Junpeng Zhang. Resonance4D: Frequency-domain motion supervision for preset-free physical parameter learning in 4D dynamic physical scene simulation. *arXiv preprint arXiv:2604.01994*, 2026.

Tianyuan Zhang, Hong-Xing Yu, Rundi Wu, Brandon Y. Feng, Changxi Zheng, Noah Snavely, Jiajun Wu, and William T. Freeman. PhysDreamer: Physics-based interaction with 3D objects via video generation. In *European Conference on Computer Vision (ECCV)*, 2024.

Licheng Zhong, Hong-Xing Yu, Jiajun Wu, and Yunzhu Liu. Reconstruction and simulation of elastic objects with spring-mass 3D Gaussians. In *European Conference on Computer Vision (ECCV)*, 2024.

A SUPPLEMENTARY VIDEOS

The supplement includes three videos that make the gauge visible, and a combined video that concatenates all three behind title cards on one canvas (26 s), `S_all_supplementary.mp4`, which is the recommended starting point. Related still figures are Fig. 1 (S1), Fig. 5 (S2) and Fig. 20 (S3). Each is self-contained: every panel carries its material setting, panels are colour-coded by whether they sit on the gauge orbit, a live PSNR or deflection readout runs under each panel, and a caption below the moving area states what the video shows and what follows from it.

- **S1 – `S1_neuma_jellyduck_gauge.mp4`.** NeuMA's released jellyduck adaptor (NeurIPS'24) rendered under baseline, gauge twin (stress$\times 4$, $\rho \times 4$) and density-only ($\rho \times 4$), in the same three groups as Fig. 1, animated over all 151 frames: NeuMA's own rasterizer, our splat re-render, and the difference *of that splat* against the baseline at $20\times$ gain. Across the sequence the gauge twin holds 105–106 dB against the baseline in NeuMA's own renders—at most one 8-bit level, max $|\Delta| = 3.9 \times 10^{-3}$, trajectory RMS 3×10^{-8}—and 145 dB in the splat, while density-only visibly flattens under gravity (32 dB in NeuMA's renders, 17 dB in the splat). The material's absolute stiffness changes $4\times$ with no observable consequence.
- **S2 – `S2_counterfactual_2d.mp4`.** Two worlds on one gauge orbit (α=8, equal c) in a 2×2 grid: rows are the protocol (passive free vibration; a new known tip load), columns the two worlds. The passive row is pixel-identical throughout (PSNR $\to \infty$); under the known load the soft/light world deflects $8.0000\times$ further, matching α exactly. Held-out-frame PSNR cannot tell the top-row worlds apart; counterfactual force-response separates them.
- **S3 – `S3_ficus_gauge.mp4`.** PhysGaussian's released ficus 3DGS asset simulated with a standard MLS-MPM (the continuum method class PhysGaussian uses; not its released code), shoved by an initial velocity—kinematic forcing, the gauge-*preserving* case of Prop. 3—and shown as baseline $\|$ gauge twin $(5E, 5\rho)$ $\|$ a different-c world $(2E, \rho)$. The gauge twin is indistinguishable (trajectory RMS 2.6×10^{-15}; 261–268 dB per frame) while the different-c world separates to 22–25 dB. (The audit figure, Fig. 20, is a separate, lower-resolution run of the same comparison.) One camera serves all three panels, its extent fitted over every frame of all three worlds. The absolute modulus is gauge-arbitrary on the real asset.

What "PSNR against the baseline" means here. It is $10 \log_{10}(\mathrm{MAX}^2/\mathrm{MSE})$ with $\mathrm{MAX} = 1$ and pixel values in $[0, 1]$, computed between the frame of one material setting and the frame of the *baseline* setting at the same time index, the same camera and the same intensity normalisation, over all pixels and colour channels. It is a similarity between two candidate worlds, not reconstruction quality against a ground-truth image: the reference is itself a simulation, and the question is whether any observer could tell the two apart. For the supplementary videos we give the range of per-frame values; elsewhere a single value refers to the frame named or, if none is named, to the whole stacked sequence.

When the PSNR is infinite, and when it is not. PSNR is $10 \log_{10}(1/\mathrm{MSE})$, so it diverges only when two images are *bit-identical*. That happens in our own float64 modal solver, where a gauge factor that is a power of two leaves E/ρ exactly unchanged and the rendered frames agree to the last bit (Fig. 5). It does *not* happen for any solver carrying round-off: NeuMA is float32 and its gauge-twin trajectories differ by RMS 3×10^{-8}, which reaches its 8-bit render as at most one level (106 dB); our MLS-MPM ficus run differs by 2.6×10^{-15} (261 dB). We report the measured value in each case. An earlier version of our analysis code returned ∞ whenever the MSE fell below 10^{-12}, which manufactures an infinity that is not there; the released code returns ∞ only for an exactly zero MSE.

B PROOFS AND GAUGE CLASSIFICATION

Proof of Proposition 1. $K_0\phi_j = \omega_j^2 M_0 \phi_j$ gives $\omega_j^2 = (E/\rho)\lambda_j$, with λ_j the j-th eigenvalue of the pencil, both it and ϕ_j depending only on geometry, boundary conditions and the fixed dimensionless parameters. g_α preserves E/ρ, hence every ω_j and ϕ_j; with identical initial conditions the superposi-

tion $u(t) = \sum_j \phi_j \, (a_j \cos \omega_j t + b_j \sin \omega_j t)$, whose coefficients are fixed by the initial displacement and velocity, is identical. □

Proof of Proposition 2. $\tau \to \alpha\tau$ and $\rho \to \alpha\rho$ give a unchanged. In an MLS-MPM transfer the grid velocity is a ratio whose numerator ($p_{\text{mass}} v + \text{affine} \cdot dx$, with p_{mass} the particle mass, C its affine velocity matrix and affine $=$ stress $+ p_{\text{mass}} C$) and denominator (the transferred mass) both scale by α, so the ratio is unchanged. A degree-one threshold $\|\text{dev}\, \tau\| \le f(\theta_\sigma)$ on the deviatoric stress satisfies $f(\alpha\theta_\sigma) = \alpha f(\theta_\sigma)$, so the projection commutes with the scaling and the plastic update is identical. □

What gravity does fix. Corollary 1 says gravity does not remove the (E, ρ) gauge; it removes a different one. Since g has dimensions L/T^2, a known $g = 9.81\,\text{m/s}^2$ and gravity-driven motion pin the combination L/T^2—one generator's worth of information: with a known frame rate it yields the metric scale, with a known object size the time scale, and never E.

Reading Table 3. The middle column is a theorem (Proposition 2); the right column is not its converse. Another symmetry, a compensating parameter change, or a protocol that fails to excite a parameter can each leave a g_α-invariant quantity unidentified, which is why we report observability under the stated protocol rather than asserting identifiability. The surviving combinations E/ρ, κ/ρ and μ_{visc}/ρ are dimensional: without a known metric scale and frame rate (Table 6) the length and time generators reduce them further, leaving only the dimensionless groups and the transit number cT/L, with T a time scale.

Table 3: **What the stress–density scaling removes, and what it leaves.** The middle column is a *theorem*: these parameters are moved by g_α and only the listed combination survives. The right column is *not* its converse—surviving g_α is necessary for identifiability, not sufficient. The surviving combinations are *dimensional*, recoverable in physical units only given a known metric scale and frame rate. EOS = equation of state. *thm* = Prop. 2; *sim* = measured in our MLS-MPM solver; *ext* = measured in a released third-party solver.

Parameter	appears in	moved by g_α	survives only as	evidence
E, Lamé λ, μ	all solids	✓	$E/\rho = c^2$	thm, sim, ext
ρ	all	✓	—	thm, sim, ext
yield stress σ_Y	von Mises plastic	✓	σ_Y/E, the yield *strain*	thm, sim, ext
cohesion c_{coh}	Drucker–Prager	✓	c_{coh}/E	thm *only*
bulk modulus κ	fluid	✓	κ/ρ	thm, sim
viscosity μ_{visc}	fluid, viscoelastic	✓	μ_{visc}/ρ, *kinematic*	thm, sim
Poisson ratio ν	all solids	✗	itself (dimensionless)	thm, sim
friction angle φ	granular	✗	itself (dimensionless)	thm, sim, ext
EOS exponent γ	weakly-compr. fluid	✗	itself (dimensionless)	thm, sim

Relation to dimensional analysis. Table 3 is a Buckingham-Π statement (Buckingham, 1914), and dimensional similarity has recently been recast as a gauge theory (Muntz, 2026); what is added here is the instantiation for the constitutive blocks current systems optimize and its verification in released solvers. The similarity group has further generators (length and time), and Table 6 states per protocol which each removes.

Identifiability is a property of the (physics, protocol) pair. Gauge membership is not protocol-dependent, but *observability* of the non-gauge parameters is, and by a wide margin. The friction angle—which Table 3 marks identifiable—moves the observable by 0.0023% under a sticky-floor shear-drop but 0.72% under the canonical granular column collapse, a $319\times$ swing from the protocol alone, with runout distance changing 17% under a $1.5\times$ friction change. A method reporting a friction angle from data that never mobilizes friction is in the same position as one reporting a modulus from passive video, for a different reason. We should be clear about the scope of the granular row: Drucker–Prager with a single friction angle is the block the benchmarks implement, not a claim about granular media in general. Rate-dependent $\mu(I)$ rheologies (Jop et al., 2006) introduce an inertial

number and further dimensionless groups, and the protocol-dependence we measure here is long familiar in that literature.

Counterfactual distinguishability (special case of Prop. 3).

Proposition 4 (Counterfactual distinguishability)**.** *Apply a known external force F in absolute units (not $\propto \rho$), producing deflection $\propto F/E$. For two gauge-equivalent members (E, ρ) and $(\alpha E, \alpha\rho)$ the ratio of responses is $1/\alpha \neq 1$. Members of one passive equivalence class are therefore* counterfactually distinguishable *under a known force while indistinguishable in passive video.*

B.1 The compound gauge

The orbit of Prop. 1 is one-dimensional, but it is not the whole of what a video cannot see. Perspective projection is invariant to simultaneous scaling of scene coordinates and camera translations, since it depends on ratios such as X/Z; a rig with a metrically calibrated baseline already breaks this, but monocular capture and uncalibrated structure-from-motion do not. Writing the dynamics in box coordinates $\hat{x} = x/L$ with L the unknown physical size of the scene gives $\ddot{\hat{x}} = (\rho L^2)^{-1}\nabla\cdot\sigma + g/L$: the physical parameters (E, ρ, L) enter only through $E/(\rho L^2)$, and through g/L when gravity drives the motion. The unobservable subspace is therefore spanned by

$$u_1 = (1, 1, 0) \ \text{(stress–density)}, \qquad u_2 = (2, 0, 1) \ \text{(metric scale)},$$

in $(\log E, \log\rho, \log L)$, and its *dimension depends on the protocol*: two without gravity, one when a known g pins L through g/L. Two assumptions belong here rather than in a footnote. This is the null space *restricted to* (E, ρ, L) with all other quantities held fixed, not the whole of what the observation cannot determine. And it presupposes a *calibrated timebase*: if the frame interval is unknown the relevant combinations are $ET^2/(\rho L^2)$ and gT^2/L, and gravity no longer recovers L on its own—length and time retain a joint ambiguity. Given the lighting-flicker artifact of Appendix E, we treat container metadata as a claim about encoding, not about physical acquisition time.

We measure this by walking along each direction (Fig. 4c). Without gravity both are flat at float64 machine zero—2.4×10^{-26} and 2.6×10^{-26} of the control—so the subspace is genuinely two-dimensional; with a known g, u_2 becomes as steep as the control ($0.66\times$) while u_1 stays at machine zero (8.4×10^{-28}).

So gravity buys back *exactly one* generator, and it is the metric scale rather than the modulus. Neither half of that is new on its own: the Π-group algebra is textbook, and gravity has been used as a metric reference in monocular vision before (Bieler et al., 2019). What the decomposition adds is the pairing—*which* generator each protocol removes. Gravity is an absolute physical quantity, which makes it tempting to suppose it anchors a modulus; it anchors length, and leaves the mass–stress generator exactly intact. That is also what a system additionally freeing a scene-scale variable is doing: searching a *larger* unobservable subspace, not a smaller one. We therefore present this as a corollary of the protocol analysis rather than an independent result.

A known mass behaves differently: it enters through $M/(\rho V)$, which is not g_α-invariant, so a balance removes u_1 when the geometry is known; if L is also unanchored the observables are $E/(\rho L^2)$ and $M/(\rho L^3)$, two constraints on three unknowns, leaving a residual generator along $(-1, -3, 1)$.

Why not a finite-difference Hessian. Along a flat direction a finite-difference Hessian reports its own truncation error, which placed these two null directions at $\sim 10^{-3}$ of the stiffest eigenvalue rather than the 10^{-26} they are. Walking the direction is immune, and a small eigenvalue is not evidence of flatness unless it falls as h^2.

C Protocol Verification and Validation Experiments

Protocol separation. Table 4 and Fig. 4 give the seven-protocol verification of Proposition 3 summarized in §4.

Gauge-blind validation, measured. We measure Corollary 2 in the metrics this literature reports. Two candidates are scored against a battery of held-out protocols: the *gauge-equivalent candidate*

Table 4: **Which protocols break the (E, ρ) gauge.** Separation of two gauge-equivalent worlds (α=4) as a fraction of the motion amplitude, in our MLS-MPM solver. The field's standard protocol—passive video, with or without gravity—sits entirely on the preserved side.

Protocol	Prop. 3	relative separation
passive free vibration	preserved	3.4×10^{-12}
gravity as a density-independent acceleration	preserved	1.1×10^{-12}
self-weight as a body force ($\propto \rho$)	preserved	1.1×10^{-12}
displacement-controlled actuation	preserved	5.3×10^{-13}
known absolute tip force	broken	0.768
calibrated pressure on a face	broken	0.772
weighed added mass at the tip	broken	0.598

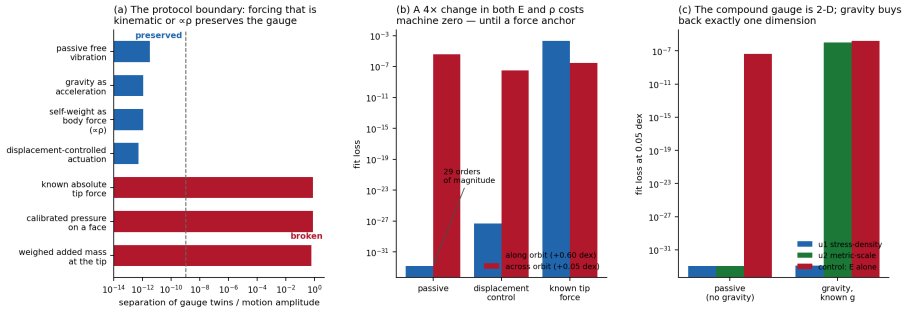

Figure 4: **The boundary, the null space, and its dimension.** (a) Separation of two gauge-equivalent worlds under seven protocols: kinematic and density-proportional forcing preserve the gauge at machine precision, an absolute force or mass breaks it, and nothing lands between. (b) Loss along the orbit (+0.60 dex, a 4× change in *both* E and ρ) against loss across it (+0.05 dex): 29 orders of magnitude apart under passive observation and under displacement control, and reversed once a force anchor is present. (c) Walking the two candidate null directions: both flat without gravity (a two-dimensional unobservable subspace), only the stress–density one flat with a known g.

$(\alpha E, \alpha\rho)$ with α=4, which is a four-fold wrong absolute modulus, and an *off-orbit control* that changes density alone. The control must fail everywhere or the metrics are degenerate and the null result means nothing. PSNR values above 258 dB in Table 5 are the numerical floor of this rendering pipeline, not a quality measurement.

Table 5: Held-out scores for a four-fold wrong absolute modulus. The first four protocols preserve the gauge—including a *new kinematic interaction*, the evidence practitioners treat as strongest—and score perfectly. Only a held-out protocol that crosses the boundary detects the error.

held-out protocol	gauge	RMSE	IoU	PSNR (dB)
gauge-equivalent candidate $(\alpha E, \alpha\rho)$				
train protocol (free vibration)	preserved	2.8×10^{-15}	1.0000	273.9
held-out frames, same protocol	preserved	6.1×10^{-15}	1.0000	267.5
held-out interaction: gravity on	preserved	1.9×10^{-14}	1.0000	258.1
held-out interaction: new kinematic drive	preserved	1.9×10^{-15}	1.0000	281.0
held-out interaction: weighed tip mass	**broken**	1.1×10^{-1}	0.0723	18.5
held-out interaction: known tip force	**broken**	7.4×10^{-2}	0.0305	15.9
off-orbit control $(E, \alpha\rho)$: fails everywhere, as it must				
best case over all six protocols	—	5.0×10^{-3}	0.4957	30.4

Counterfactual force response and intervention design. By Proposition 4, two settings on one orbit render identically on every frame, so a photometric score certifies the equivalence class and never the absolute: two worlds at α=8 with identical c are pixel-identical passively yet separate by

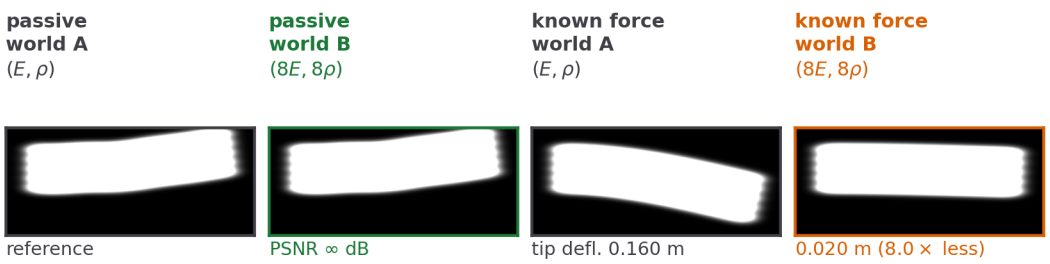

Figure 5: **Counterfactual force response is the evaluation that works.** Two worlds on one gauge orbit, $8\times$ apart in both E and ρ and so equal in elastic speed scale, through the same differentiable splat renderer. Under passive free vibration they are pixel-identical, so no held-out frame, view or sequence separates them. Under one known tip load the soft world deflects $8.0000\times$ further—exactly α—because static deflection goes as F/E and is independent of density. Supplementary video S2.

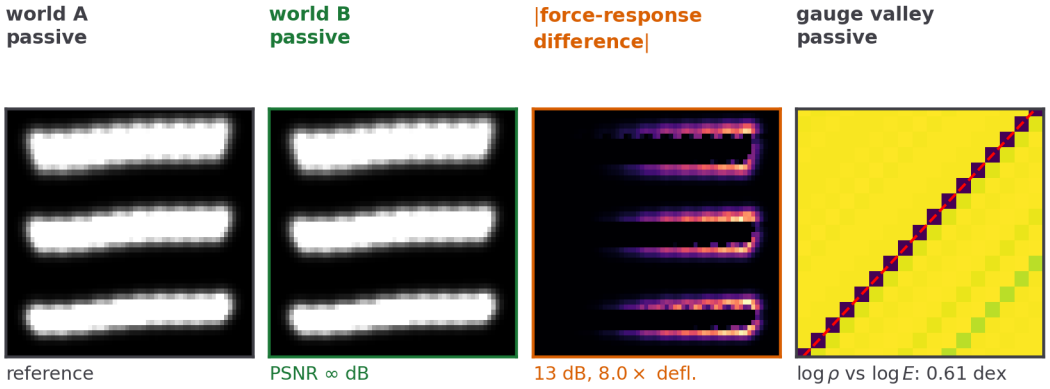

Figure 6: **The same conclusion under 3D multi-view Gaussian-splat observation.** Two cameras, a 90-node beam, isotropic kernels without opacity compositing or view-dependent shading, so we do not claim a full 3DGS renderer. Two worlds on one gauge orbit (α=8, identical c) are pixel-identical passively; the |difference| under a new known force applied by *dynamic* forced-EOM simulation is 13.2 dB with an $8.0\times$ peak-deflection ratio. The right panel is the loss over $(\log_{10}\rho, \log_{10}E)$ across ±1 dex, flat along the orbit (dashed red) at 0.61 dex of E spread. Fig. 5 is the 2D version.

9.9 dB under a new force, with exactly an $8\times$ deflection ratio (Figs. 5 and 6). Passive video carries *exactly zero* Fisher information about the absolute scale, for one frame and for all T=14 alike, while a single known tip load injects 8.3×10^4 units of it, and an information-selected load beats a poorly placed one by $73\times$ in $\sigma_{\log E}$ (Figs. 7 and 8).

The gauge valley and its anchors. The valley is identical in a 1D pilot and in the 3D splat pipeline: passive observation leaves absolute E at the width set by the prior, while a known density, a known mass and scale, or a known load each collapse it to within one grid cell (Figs. 9, 10 and 11; Table 6). Fig. 7 gives the continuous posterior widths, where the placement of the load matters.

Estimator and the Cramér–Rao bound. We infer $\log_{10} c$ by profile likelihood under an i.i.d. point-track (or pixel) noise model, locating the mode by a global grid search (the c-from-dynamics likelihood is multimodal due to frequency aliasing) followed by a local Gauss–Newton refinement; the posterior width is the Cramér–Rao bound from the curvature at the mode. Absolute E is reported as the induced interval along the gauge orbit, which collapses only when an anchor (Table 6) is supplied.

Observation-model and constitutive robustness. The gauge valley is identical (E uncertainty 0.72 dex, c 0.0 dex) across a Gaussian-splat, a wider-kernel splat, and raw point-track observations, and persists under Rayleigh-damped viscoelasticity—confirming it is a property of the physics and

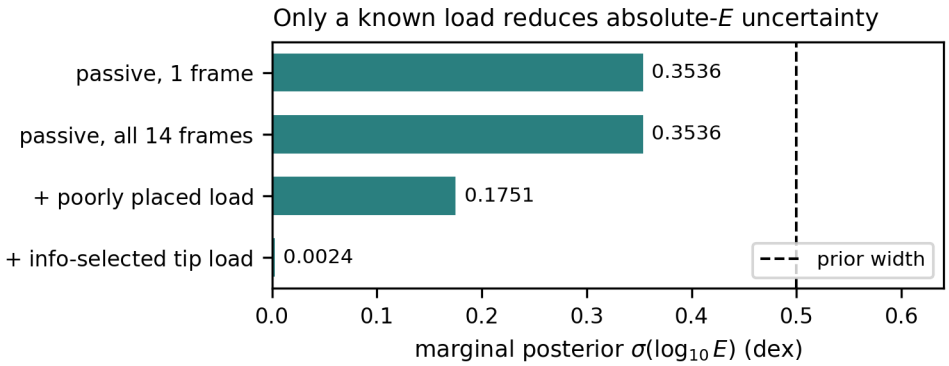

Figure 7: **Absolute-modulus uncertainty collapses only under an intervention.** Marginal posterior width on $\log_{10} E$ per protocol, with 0.5 dex priors on $\log_{10} E$ and $\log_{10} \rho$. Passive observation pins c and adds nothing along the orbit, leaving the prior-implied width $0.5/\sqrt{2} = 0.354$ dex whatever the frame count; an information-selected tip load beats a poorly placed near-clamp one by $73\times$ (0.0024 against 0.1751 dex). *Which* intervention matters as much as whether.

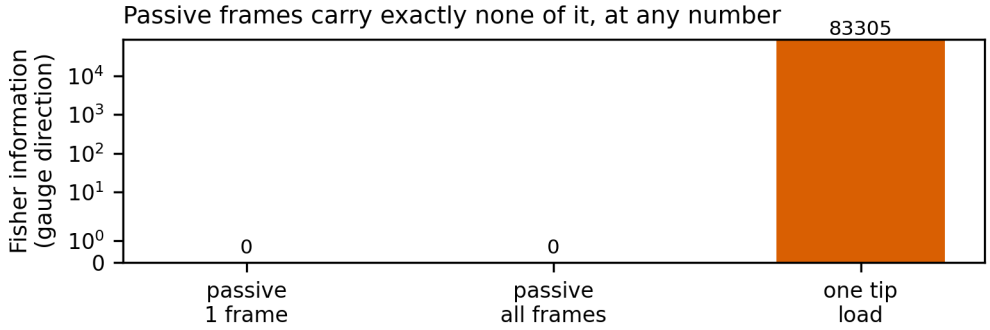

Figure 8: **One poke beats a thousand passive frames.** Fisher information along the gauge direction: exactly 0 for one passive frame and for all 14, and 8.3×10^4 for a single known tip load. The zero is algebraic, not a small number—no quantity of passive video carries any of it.

observation protocol, not of any particular kernel. (We do not claim a NeRF volume renderer; the widened kernel tests robustness to the observation kernel.)

Calibrated posterior over the invariant. Our posterior over c passes rank-based simulation-based calibration (SBC) under the assumed model (χ^2 p=0.15; coverage $0.48/0.89/0.94$ at nominal $50/90/95\%$; Fig. 12), and absolutes are reported only conditional on a declared anchor. Calibration holds only under the assumed model: a 5% error in the assumed geometry scale shifted the recovered c by 0.20 dex in our tests, since metric scale is itself a gauge direction that needs an anchor.

Coverage when a fixed density is treated as exact. A density can be a good anchor—it is often the easiest property to measure, and shear-wave elastography converts a measured wave speed to a modulus through a stated tissue density. What matters is that the anchor's uncertainty is carried. From $E = \rho c^2$,

$$\mathrm{Var}(\log E) \;=\; \mathrm{Var}(\log \rho) \;+\; 4\,\mathrm{Var}(\log c) \;+\; 4\,\mathrm{Cov}(\log \rho, \log c), \tag{1}$$

which reduces to its first two terms for a density anchor independent of $\hat{c}$—the case we assume throughout—and fixing the density sets the first of those to zero. Measured as coverage (Fig. 13), a nominal 95% interval for absolute E with ρ fixed covers 24% for water, 2% for the food pastes PAC-NeRF simulates and 0% for granular media and foams, while the *same estimator with the same c-uncertainty* carrying a class density prior covers 94% throughout. The harm also *grows* as the video improves: a tighter c shrinks the interval while the density offset stays put.

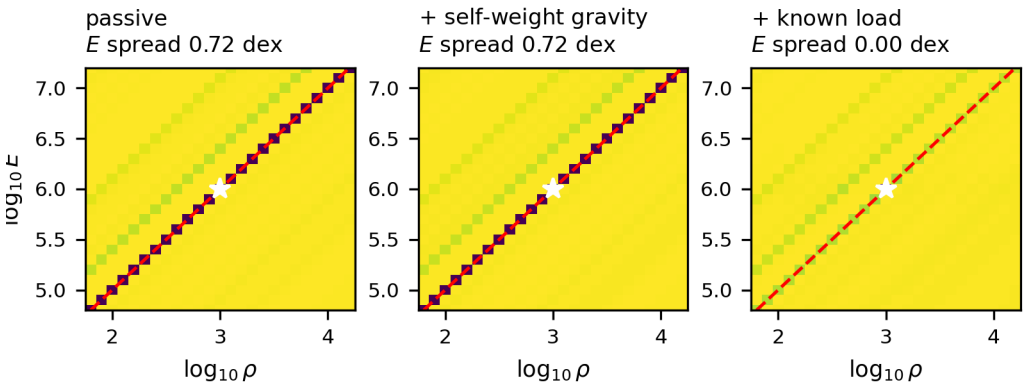

Figure 9: **The gauge valley, and what an anchor does to it.** Image-space loss over $(\log_{10}\rho, \log_{10} E)$ on a 25×25 grid spanning ± 1.2 dex, through the 2D splat observation model. Dashed red is the orbit $E = c_0^2\rho$, the white star the truth. Passive observation is flat along it and self-weight gravity leaves it flat—both at 0.72 dex of E spread—while one known load collapses the valley to a point (0.00 dex).

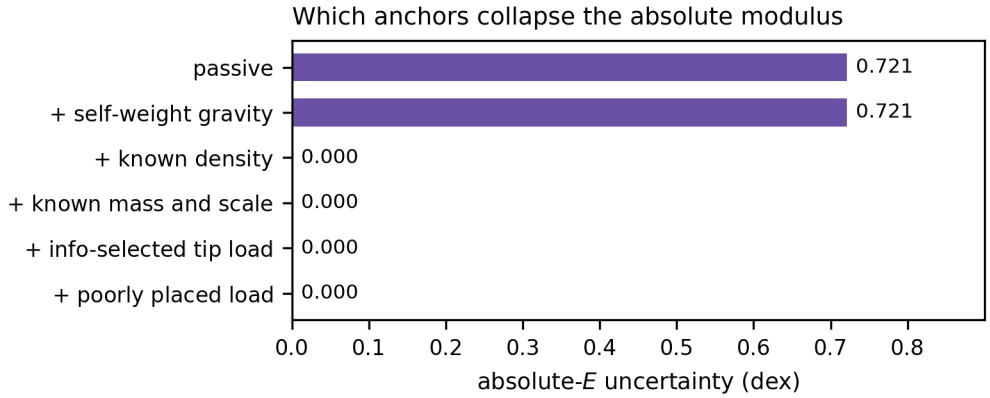

Figure 10: **Which anchors collapse the absolute modulus.** Absolute-E uncertainty per protocol. Passive observation and self-weight gravity both stay at the width set by the prior box (0.721 dex); a known density, a known mass with scale, or a known load each collapse it to 0.000 dex. Spreads are measured on the likelihood grid, so any calibrated load collapses to one cell; Fig. 7 gives the continuous widths (0.0024 against 0.1751 dex for a well and a poorly placed load). Gravity sits on the preserved side, the entry practitioners find least intuitive.

Class density priors checked against measured specimens. GAUGE (Wang et al., 2026a) publishes a mass, a metric density, a Poisson ratio and a Young's modulus obtained by mechanical testing with digital image correlation for each of its deformable specimens, so the anchor error is measurable rather than assumed. Its six distinct characterised specimens span $\rho = 12.8$ to $1298.7\,\mathrm{kg/m^3}$. For its stick-stack specimens ($\rho \approx 1280$) a hard-coded ρ=1000 misstates E by only $0.8\times$—the practice is essentially correct there. For its foams ($\rho = 12.8$–27.2) the same constant misstates E by $37\times$ to $78\times$, an anchor error of 1.6–1.9 dex. That is larger than the 1.1 dex we assumed for the foam class in Fig. 13, so the coverage numbers above understate the effect for the one material family where an independent measurement exists. The same benchmark thus supplies both ends of the argument: a regime where fixing the density is fine, and a regime where it is wrong by nearly two orders of magnitude.

Classification across constitutive models. Figure 14 shows the verification summarized in §7: the transformation of Proposition 2 leaves all four constitutive blocks exactly invariant, while partial scalings move the object by 0.3–2% of its extent.

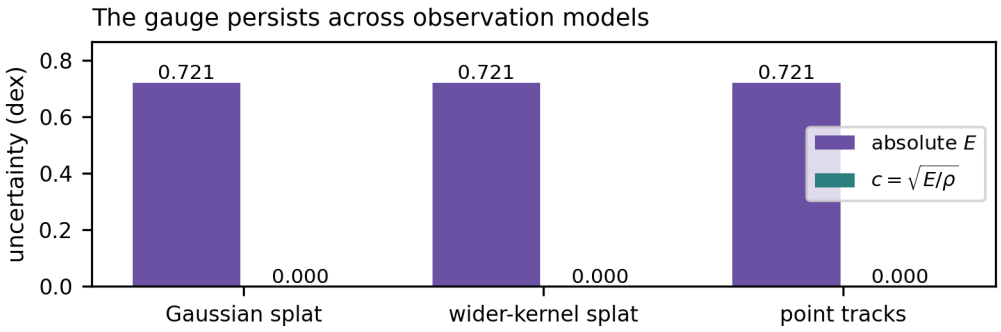

Figure 11: **The gauge belongs to the physics and protocol, not to the observation kernel.** Absolute-E and c uncertainty under three observation models—a Gaussian splat, a wider-kernel splat, and raw point tracks. All three give 0.721 dex on E and 0.000 on c, so the valley is not an artifact of any particular kernel. (We do not claim a NeRF volume renderer; the widened kernel tests robustness to the kernel.)

Table 6: Mass-scale gauge atlas. ✓: the mass-scale gauge is removed, so this gauge does not preclude recovery; ✗: it is not; (✓): removed given the row's anchor. Removal is necessary and not sufficient for identifiability, which we claim nowhere (§4). The *counterfactual* column marks protocols under which the recovered parameters suffice to predict the response to a calibrated new intervention. The last two rows differ in how the load arises: "known external load" is passive observation plus a load that happens to be calibrated, "active" is an intervention placed by design. Absolute (E, ρ) require an anchor injecting a ρ-independent force/mass/scale; passive observation—with or without self-weight gravity—identifies only c (and only given metric scale and frame rate). Evidence: *deriv* = derived analytically; *sim* = measured in our MLS-MPM solver; *OED* = optimal experimental design, i.e. the load was placed to maximise Fisher information rather than by hand.

Observation protocol	E	ρ	$c{=}\sqrt{E/\rho}$	force scale	counterfactual	evidence
Passive, unknown metric scale	✗	✗	✗	✗	✗	deriv
Passive, known scale + frame rate	✗	✗	✓	✗	✗	deriv+sim
+ self-weight gravity (g known)	✗	✗	✓	✗	✗	deriv+sim
+ known density ρ	(✓)	given	✓	✗	✓	sim
+ known mass & volume	(✓)	(✓)	✓	✗	✓	sim
+ known external load F	✓	✓	✓	given	✓	deriv+sim
Active: response to known force	✓	✓	✓	given	✓	sim (OED)
Passive, unknown frame rate	✗	✗	✗	✗	✗	deriv

Recovering the classification from a black box (`gaugefix`). The classification is also *discoverable* (Fig. 15). Our released tool takes a simulator as an opaque callable plus a dictionary of positive numbers, finite-differences the observable in log-parameter space, rationalizes candidate null directions to integer exponents, and tests each by travelling a finite distance rather than by thresholding a derivative—which matters because a non-smooth solver defeats the derivative outright and because detection is bounded by the precision the model reports, often `float32`. Knowing nothing of what its inputs mean it recovers the predicted orbit and an equivalent basis of invariants for all four blocks. We use "certify" in this numerical sense only: a finite black-box test is evidence for a scaling symmetry, not a proof of structural invariance, and on PAC-NeRF's plastic solvers it declines rather than certifying.

Estimators on the orbit. Under each method class's declared observation model the ambiguity is structural: from broad ± 1.2 dex random starts in *both* coordinates, the third of runs that reach the noise floor recover c exactly while their absolute E spans 0.51 dex and tracks the initial valley coordinate at Pearson $r{=}0.998$—the reported E is set by initialization, not by the video. (A second set of starts placed deliberately *along* the valley merely persists there, which is a property of the

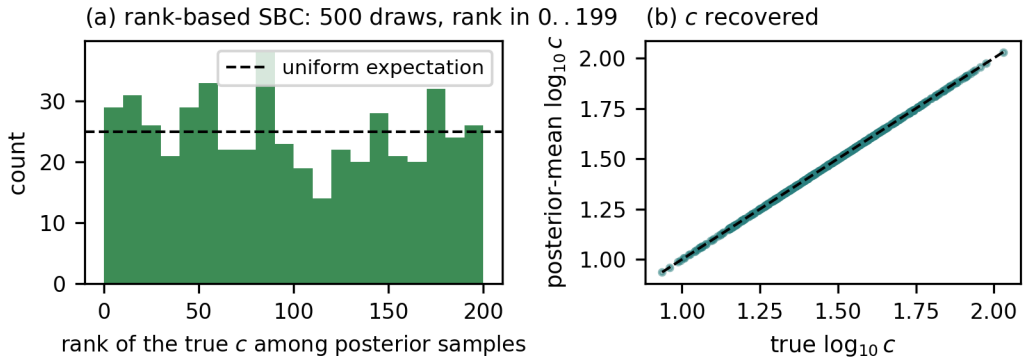

Figure 12: **The quotient posterior is calibrated even though the absolute is not.** Rank-based simulation-based calibration of the Laplace posterior over c under the assumed model: 500 draws, each ranked among 199 posterior samples. (a) The rank histogram is uniform (χ^2 p=0.15) and interval coverage is $0.48/0.89/0.94$ against the nominal $50/90/95\%$. (b) Recovery of $\log_{10} c$, mean posterior width 1.5×10^{-4} dex. Calibration of the invariant is not calibration of the absolute, which has no width at all until an anchor supplies one.

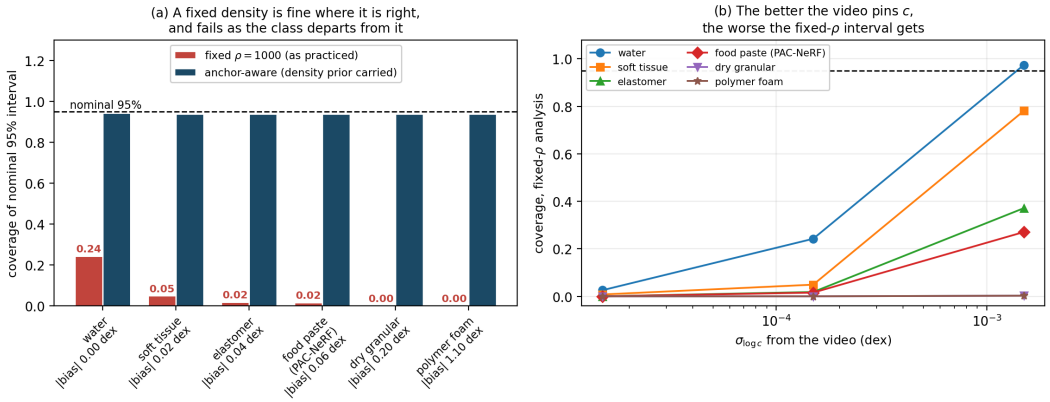

Figure 13: **An undeclared density anchor destroys interval coverage, and the damage scales with how wrong the anchor is.** (a) Coverage of a nominal 95% interval for absolute E, per material class, ordered by the offset between the assumed ρ=1000 and the class mean. Both estimators use the *same* empirically calibrated c-uncertainty, so the only difference between them is whether the anchor's uncertainty is carried; the gap is therefore attributable to the anchor alone. The practice is defensible for water and degrades monotonically, reaching 0% for granular media and foams. (b) The same quantity against the precision with which the video determines c: as c becomes better determined the fixed-density interval narrows around an offset that does not shrink, so coverage *falls*. Better vision makes the undeclared anchor worse.

initialization and not evidence about the optimizer; we report it as such.) A feed-forward regressor on a gauge line predicts the training-prior mean ($R^2 \approx 0$, capacity-invariant across $8/16/32$ widths)—and positive controls confirm the *identical* network learns c (R^2=0.99) and learns E once ρ is fixed (R^2=0.99), so the $R^2 \approx 0$ is genuine gauge-blindness, not undertraining; and an oracle given c plus a density anchor recovers E exactly. The reported absolute is set by initialization, prior, or anchor—not by the video.

Real multi-view capture. We repeat the test on real high-speed capture (Fig. 16), using the released multi-view captures of Spring-Gaus (Zhong et al., 2024) rather than a rig of our own: three calibrated cameras at 119.88 fps observing real 8–10 cm deformable objects bouncing on a table, with geometry reconstructed by visual-hull carving from the real silhouettes and initial conditions

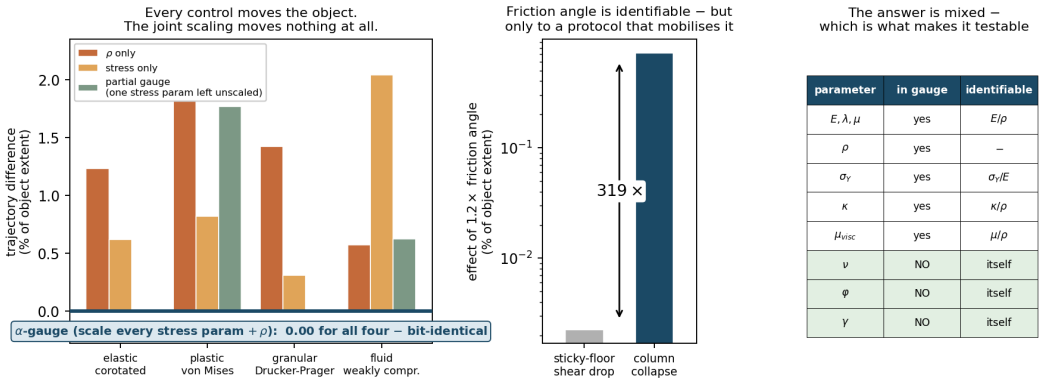

Figure 14: **The classification across the constitutive zoo.** (left) Scaling density alone, stress alone, or all-but-one stress-dimensioned parameter all move the object by 0.3–2% of its extent; scaling every stress-dimensioned parameter together with density moves it by *exactly nothing*, in all four regimes—including two nonlinear plastic return maps, where the invariance is not an identity of the setup. (middle) Gauge membership is protocol-independent, but observability is not: the friction angle, which the classification marks identifiable, is 319× more visible under a granular column collapse than under a sticky-floor shear drop — the same parameter is practically identifiable under one protocol and practically invisible under the other. (right) The table is mixed, so every tested row is a prediction that could have failed.

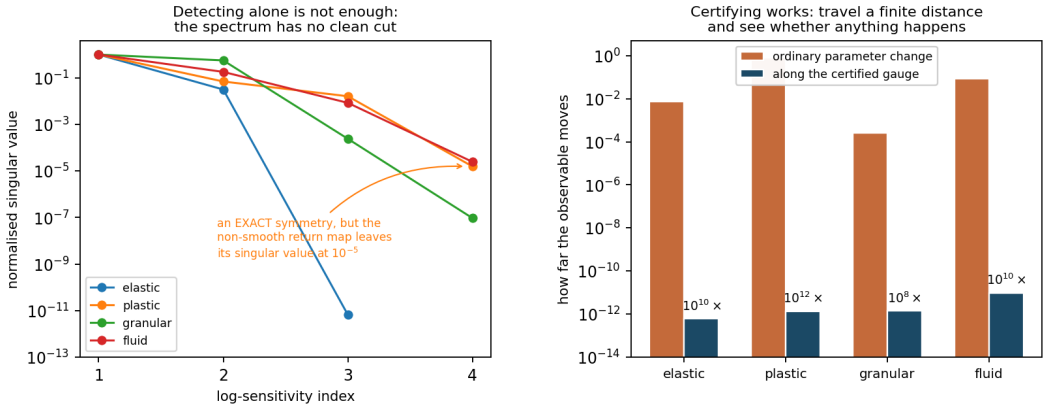

Figure 15: **Detect, then certify.** (left) A threshold on the log-sensitivity spectrum cannot deliver the classification: the plastic gauge direction is an exact symmetry, yet its non-smooth return map leaves the singular value at 10^{-5}, indistinguishable from a merely weakly-identified parameter. (right) Certification settles it. Travelling a finite distance along a candidate direction moves the observable 10^{8}–$10^{12}\times$ less than an ordinary parameter change of the same size, a gap no threshold on derivatives has to guess at.

read off the triangulated centroid. The pipeline validates against a known constant: fitting the pre-impact free-flight acceleration of the reconstructed volume centroid gives $11.0 \pm 1.4\,\mathrm{m/s^2}$ across the five objects (mean $\pm$ s.e.m., 0.85σ from g=9.81), with sub-millimetre fit residuals (0.14–0.43 mm), which checks calibration, triangulation, reconstruction and frame rate together. The per-object scatter (7.3–$15.4\,\mathrm{m/s^2}$) is a precision limit rather than a defect: the released masks give only 6–7 pre-impact frames, and over that $\sim$50 ms window gravity contributes $\sim$1.2 cm of curvature on top of $\sim$18 cm of linear fall. Nothing below depends on this check: every result in this paragraph holds identically whatever it returns. Pushing two gauge-equivalent materials (α=5) through the *real* cameras and scoring against the *real* masks gives a silhouette-IoU difference of *exactly zero* at every frame, on every one of the five objects (trajectories agree to 7×10^{-12}), while a different c differs by 0.19–0.55 IoU. The posterior over the real masks is flat along the orbit and sharp across it. We score with a

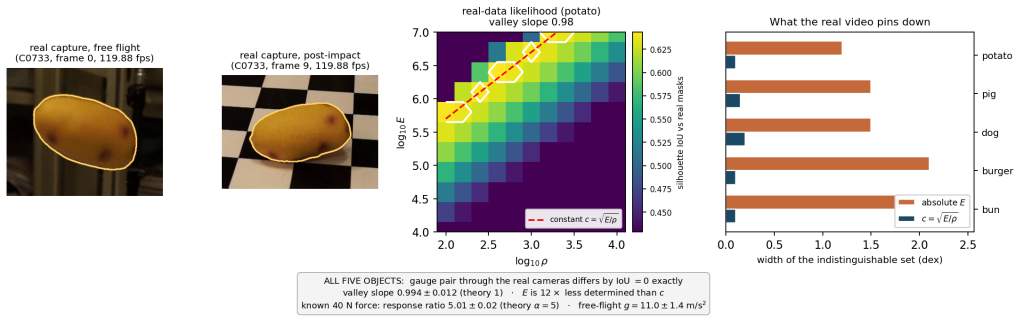

Figure 16: **The gauge on real captured video.** Real 8–10 cm objects dropped onto a table, recorded by three calibrated cameras at 119.88 fps; geometry is carved from the real silhouettes (outline). (right-centre) The likelihood over $(\log E, \log \rho)$ scored against the *real* masks is a valley along $c{=}\sqrt{E/\rho}$. (right) On every object the real video leaves absolute E uncertain across more than a decade while pinning c, and two materials related by the scaling are *exactly* indistinguishable through this apparatus. The known-force separation is a *counterfactual simulation* through the same calibrated cameras—no force was applied to the recorded objects—and it recovers precisely the predicted factor.

stated observation model rather than a threshold: each silhouette pixel is a Bernoulli trial whose flip rate is *calibrated per object* from the reconstruction's own reprojection disagreement ($\varepsilon = 0.037$–0.043), with the sample size divided by the measured spatial correlation patch of the residual, and we grid $(\log c, \log \rho)$ rather than $(\log E, \log \rho)$—in the latter the resolution in c is fixed by the *ratio* of the two step sizes, so refining the grid never refines c. Under a flat prior over a declared 2-dex box in density, the marginal for $\log E$ *fills that box on all five objects*: the data supply no information about the absolute scale at all. The marginal for $\log c$ is narrower than one grid cell, < 0.03 dex. The wave speed is therefore determined at least $67\times$ more tightly than the modulus—a bound, not a tuned number. (The "$12\times$" in Fig. 16 comes from the earlier $(\log E, \log \rho)$ grid, whose resolution in c is limited by its step sizes.) We report c as a bound rather than a width deliberately: the residual at the optimum (0.89–0.95 pixel agreement against a 0.96 floor) is systematic model error, not the i.i.d. pixel noise the likelihood assumes, so a nominal width for c would be exactly the false precision this paper argues against. What survives that objection is the contrast between the two directions, and it is large. A new known 40 N force separates the same pair by a response ratio of 5.01 ± 0.02, against a predicted $\alpha = 5$. Our forward model is a poor model of a tumbling plush toy—it sits at the reconstruction noise floor in free flight (mean IoU 0.83 against a 0.85 reprojection ceiling) and degrades sharply after impact, to 0.49 over the window the likelihood actually scores—and this matters not at all for the claim: the two simulations agree to the last bit whether or not either matches the toy, which is what it means for the ambiguity to be a property of the observation model rather than of fit quality. Accordingly we report the recovered c as model-conditional and do not claim it as a measurement of the object.

D Released-System Audit Details

Table 7 summarizes the audited systems and the mode of each audit; the paragraphs below give the details system by system.

PAC-NeRF (ICLR'23), source audit of the canonical benchmark. PAC-NeRF is the reference continuum system-identification benchmark, and its simulator implements exactly the four constitutive blocks of Table 3. Two facts follow from its released source. First, `train.py` builds the physical-parameter optimizer solely from the keys of `physical_params`, and density appears in *no* configuration's `physical_params`: it is never an optimization variable in any of the 25 configurations, and where a literal is present (20 of 25; the other five inherit it through `_base_`) that literal is `rho = 1000`. Density is therefore never estimated; it is the gauge choice. Second, its return maps satisfy the hypotheses of Prop. 2 verbatim: the von Mises test is $\|2\mu\hat{\varepsilon}\| \geq \sqrt{2/3}\,\sigma_Y$,

Table 7: **Where the density dependence could be executed or evaluated analytically, the measured sensitivity matches the predicted gauge dependence; the source-only rows identify the gauge-fixing mechanism instead.** The middle column is that measured log–log sensitivity, where we could measure it; Prop. 1 predicts exactly 1 when the gauge-fixing quantity is a density. Vid2Sim's 0.00 is *not* identifiability—it never reads a density, so its absolute is fixed by the training prior instead. VVT is the case done right: the same sensitivity of 1, but to a prior it *declares*. M-PhyGs is from an independent lineage: density is optimized and the scale is still set outside the data. For NeuMA, which learns a constitutive law rather than a modulus, $\hat{E}$ is the absolute stiffness scale of that law. *Mode*: *code* = we ran the method's own released solver; *asset* = we ran its released asset through a standard MLS-MPM; *ckpt* = we ran its released checkpoint; *source* = source audit with parameter perturbation.

Method	Gauge fixed by a	$d\log\hat{E}/d\log a$	Audit evidence	Mode
PAC-NeRF (Li et al., 2023)	`rho`, a constant	1.003 ± 0.006	re-fit over $16\times$ ρ	code
PhysGaussian (Xie et al., 2024)	hand-set ρ	1.00	259 dB at equal c	asset
NeuMA (Cao et al., 2024)	ρ=1000	1 (exact)	gauge render RMS 3×10^{-8}	ckpt
VVT (Feng et al., 2022)	*declared* prior mean	1.000 (prior)	estimator linear in the prior	code
Vid2Sim (Chen et al., 2025)	prior + time scale	0.00 (ρ)	density-blind; $8\times$ re-timing	ckpt
GIC (Cai et al., 2024)	no ρ optimizer step	—	ρ gradient built, discarded	source
M-PhyGs (Wada et al., 2026)	box, init., optimizer	—	E/ρ at float32 ε	source

i.e. yielding at the *dimensionless* strain $\sqrt{2/3}\,\sigma_Y/2\mu$; and the Drucker–Prager increment carries the coefficient $(d\lambda + 2\mu)/(2\mu) \cdot \alpha_{\text{friction}}$, which depends only on ν and the friction angle.

Reading the two together gives a per-scene verdict, and it is exactly the mixed one Table 3 predicts. The 14 elastic scenes estimate (E, ν): ν is identifiable, E is not. The 3 fluid scenes estimate (κ, μ), and 3 of the 6 von Mises scenes estimate $(\kappa, \mu,$ plastic viscosity, $\sigma_Y)$—in these 6 scenes *every* estimated quantity lies in the gauge, so the entire reported parameter vector is set by the hard-coded density rather than by the video. The other three von Mises scenes estimate (E, ν, σ_Y), of which only ν survives. Running their released simulator confirms the derivation rather than only the reading of it. Applying the predicted scaling leaves the trajectory *bit-identical* (0.00) in all three of their solvers we could drive—elastic, von Mises and Drucker–Prager—whereas scaling μ alone moves the elastic solver's object by 0.53% of its extent. On that solver our detector also recovers the (μ, λ, ρ) direction from the black box unaided. Their fourth material, the viscous fluid, we could not test this way: its particle count varies between frames, so trajectories are not directly comparable without tracking surviving particles. One caveat is worth recording, since it is a property of their discretization rather than of the physics: the invariance is exact when the parameters are scaled *up* ($\times2$, $\times4$, $\times16$ all give 0.00) but departs by $\sim5 \times 10^{-7}$ when scaled down, growing as the scale shrinks—the one-sided signature of an absolute threshold inside an otherwise scale-covariant solver. The granular scenes are the control that makes the classification falsifiable: they estimate the friction angle *alone*, which is dimensionless, untouched by this gauge, and strongly observable under the protocol they use—so within that one-parameter block those numbers are measurements, and our account says so.

PAC-NeRF, changing only the assumed density (Fig. 17). Source reading is not causal evidence about the estimator, so we re-ran their fit many times, holding the observations fixed and changing nothing but the density constant. We drive their released `MPMSimulator` through their `DynamicObserver`, with their (E, ν) parameterization, their sticky ground collider, and their published per-scene ground truth; the observations are their own released ground-truth particle sequences. Giving the fit *3D particle positions* instead of rendered video is deliberately generous: it removes the renderer and every photometric confound, so whatever we see cannot be blamed on rendering.

Across four elastic scenes and a $16\times$ range of assumed density ($\rho = 250$ to 4000), the recovered modulus tracks the assumption with slope $d\log\hat{E}/d\log\rho = 1.003 \pm 0.006$ (theory: exactly 1; largest single-scene deviation 0.013), while the attained fit changes by 0.6–4.7% and the implied invariant $\hat{E}/\rho$ by 1.7–7.3%. A practitioner who assumes water and one who assumes a dense sand will report moduli differing by a factor of four, with the same data, the same solver, and—to within a few

percent—the same residual. Neither has any way to tell from the fit which is right, and ρ appears in no configuration's `physical_params`, so the choice is never surfaced.

Two caveats, both ours rather than theirs. Our reproduction of their pipeline recovers their published modulus to within 0.07–0.44 dex, with a residual trajectory mismatch of 4–6% of object extent, because their particles are sampled from a trained NeRF geometry we do not rebuild; the *slope* this experiment measures is a ratio and is unaffected by any constant offset. And their elastic sequences contain no information about E at all until contact: stripping rigid translation from the released ground truth leaves *exactly* zero deformation for the first four to five of sixteen frames in every scene, which is what their own `hit_frame` declares. A fit scored only before contact is optimizing a flat likelihood.

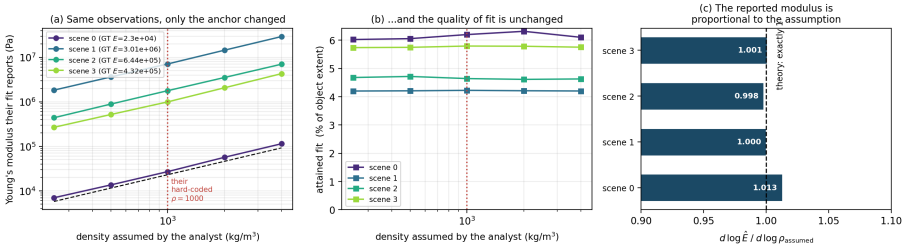

Figure 17: **The reported modulus is proportional to the density the analyst assumed.** We re-run PAC-NeRF's own fit on its own released ground-truth sequences, changing nothing but the density constant. (a) Recovered $\hat{E}$ against assumed ρ, four scenes, log–log; the dashed line is the theoretical slope of 1 and the dotted vertical marks the ρ=1000 their configurations hard-code. (b) The attained fit over the same 16× range—flat to within a few percent, so nothing in the data prefers one anchor over another. (c) Per-scene slopes against the predicted value of exactly 1.

NeuMA (NeurIPS'24), released adaptor (Figs. 1 and 18). NeuMA learns a constitutive correction on a base law (Ma et al., 2023). We load the released jellyduck adaptor and multiply the stress returned by its learned constitutive law, and its density, by four. The simulated trajectories agree to RMS 3×10^{-8} (float32), and the rendered frames agree with the baseline to 106 dB in NeuMA's own rasterizer and 145 dB in our splat re-render; density alone ×4 gives 32 and 17 dB. The absolute stiffness of its learned material is therefore fixed by the assumed density ($\rho = 1000$), not by the video.

Vid2Sim (CVPR'25), live checkpoint (Fig. 19). Vid2Sim regresses absolute E feed-forward with ρ fixed to 1000; we verified its decoding ($\hat{E} = 10^{\text{head}}$; the head is trained against $\log_{10} E$ unnormalized). (i) *Density blindness*: at fixed c, inputs with true E differing 8× (ρ=1000 vs 8000) yield *identical* $\hat{E}$ ($\hat{E}$ spread 0.0 dex vs true-E spread 0.90 dex; $R^2(\hat{E}, E_{\text{true}}) = -2.2$)—unavoidable, since the videos are pixel-identical. On this out-of-distribution asset $\hat{E}$ is in fact near-constant, so it recovers neither absolute E nor c. (ii) *Time-scale gauge* (the sharper, in-distribution test): on its own test objects (sane, varied predictions, 0.48 dex spread—confirming correct loading), re-timing the same motion changes $\hat{E}$ by 8× (0.9 dex, slope −1.49 in $\log \hat{E}$ vs $\log$ span). We report this as *sensitivity* to the assumed frame rate (the sign is opposite the naive $1/dt^2$ frequency expectation, so we do not claim a calibrated time-scale law). Absolute E is thus gauge-fixed by an assumed density *and* an assumed time scale.

PhysGaussian (CVPR'24), released asset (Fig. 20). On PhysGaussian's released ficus 3DGS asset, simulated with a standard MLS-MPM (the continuum method class PhysGaussian uses; we do not run its released code, and gauge invariance is a property of that shared continuum), equal-c settings (E, ρ) vs $(5E, 5\rho)$ render indistinguishably (258.9 dB in the audit run of Fig. 20, the float64 floor of that pipeline) while a different-c setting is distinguishable (29.5 dB); free-ρ recovery is degenerate (E uncertainty 0.57 dex, c 0.0 dex). A known-force anchor collapses E to 0.0 dex.

GIC (NeurIPS'24 Oral), code audit. GIC estimates E and ν following PAC-NeRF, and the "density fields" it builds are geometric fields used to sample particles, not mass density. Its released



Figure 18: **Audit 1, NeuMA's released adaptor: a *parameter estimator*.** Baseline and gauge-transformed (stress $\times 4$, $\rho \times 4$) renders through NeuMA's own rasterizer agree to 106 dB at the final frame (trajectory RMS 3×10^{-8}), while a density-only change drops to 32 dB. Its absolute stiffness is fixed by its ρ assumption, not by the video. Supplementary video S1 animates the full sequence.

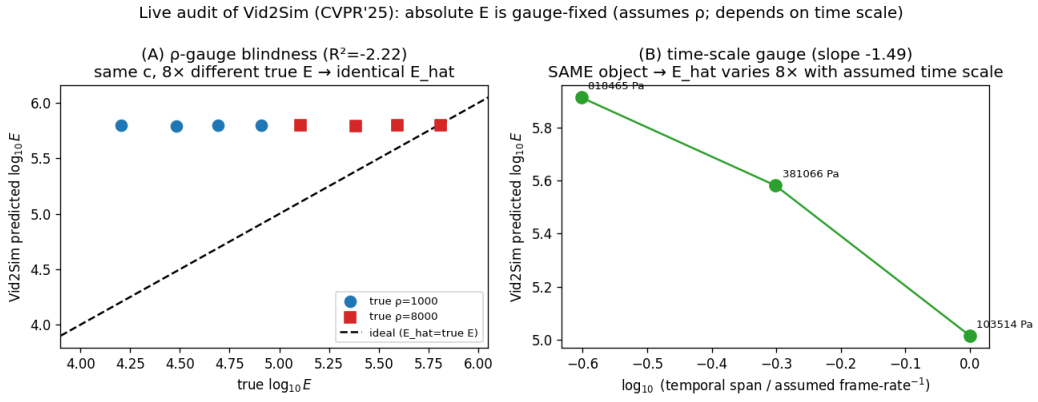

Figure 19: **Audit 3, Vid2Sim's checkpoint: a *feed-forward regressor*.** $\hat{E}$ is density-blind, and shifts $8\times$ under an in-distribution change of time scale, so its absolute is fixed by the training prior and the assumed frame rate rather than by the mechanics.

code constructs a learnable `global_rho` but never adds it to the optimizer, so ρ is held fixed and the absolute E is gauge-fixed. Supplying the missing branch takes one `elif` and a log-parameterization mirroring the one used for E; the runtime trace before and after is in §6.

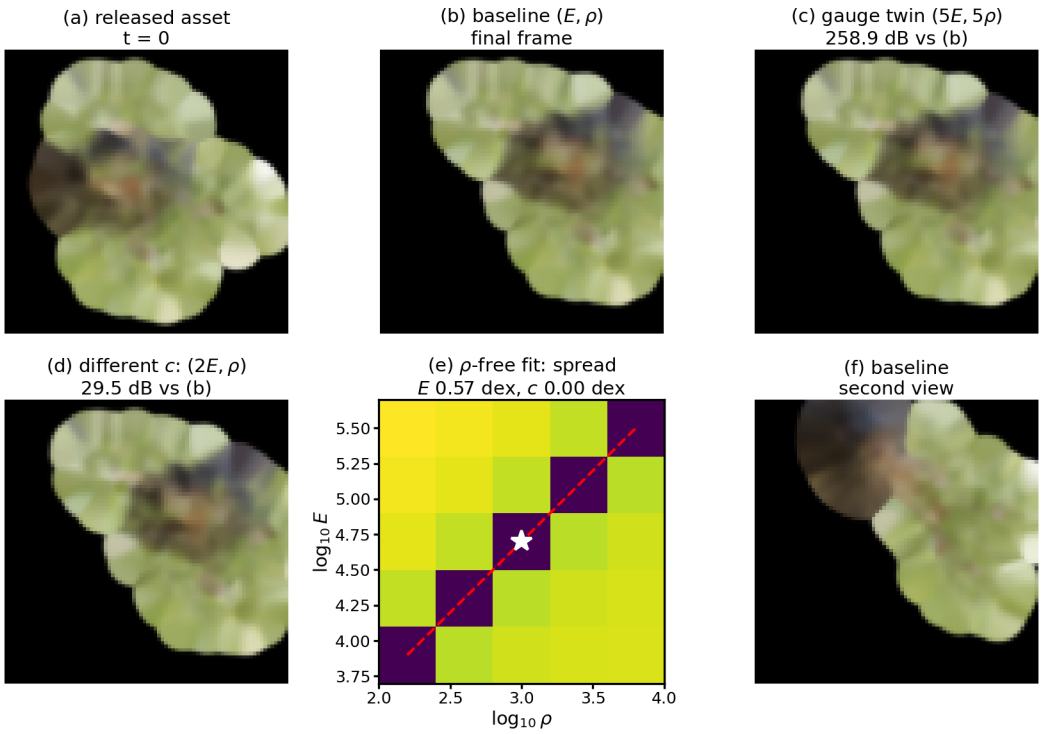

Figure 20: **Audit 2, PhysGaussian's released ficus asset run through a standard MLS-MPM: a *symmetry demonstration* on a real asset, not an audit of an estimator**, since PhysGaussian hand-sets its material values. (c) The gauge twin agrees with the baseline to 258.9 dB, the round-off floor of this run, while (d) a different c gives 29.5 dB; (e) the ρ-free fit follows the valley $E{=}c^2\rho$. Supplementary video S3 animates the three settings.

ProJo4D (TMLR 2026) and MonoPhysics (2026). Both inherit PAC-NeRF's convention: all 26 ProJo4D and all 4 MonoPhysics configurations use $\rho = 1000$ and optimize density in none. Re-running ProJo4D's own fit over an $8\times$ density range reproduces the PAC-NeRF result in a different codebase: slope 0.990 ($R^2{=}1.00000$) at 3.7% spread in attained fit. The optimizer branch that frees GIC's density does nothing in ProJo4D, whose `particle_rho` is declared without `needs_grad`, so its density gradient is `None` on all 80 iterations; MonoPhysics severs the graph with an explicit `.detach()` before simulation (Table 2).

M-PhyGs (CVPR'26 Findings), source audit. M-PhyGs (Wada et al., 2026) estimates E and ρ per material segment from real multi-view video. In the released repository the human interaction overwrites particle velocities before the particle-to-grid transfer, a kinematic Dirichlet datum; gravity and the estimated initial internal force both enter as `particle_mass` $\times$ acceleration, so both scale with ρ; and neither the ChArUco metric calibration nor the stiffness-free ground collider supplies a mass datum. The density gradient path is complete—`backward` returns `density_grad` from the solver's own adjoint—and density is stepped at learning rate 0.05 within the box $[10, 10^4]$. The material field is written `particle_E = particle_normalized_stiffness * particle_density`, so perturbing the density head by 0.53 dex moves ρ while E/ρ changes by 1.4×10^{-7} (float32 ε), and both material regularizers are exactly invariant under the gauge. In all 20 shipped configurations the density head's only input is a cluster-probability vector that the loader hard-assigns to one-hot, so with the shipped linear material model it is a lookup from a segment indicator: one free $(E/\rho, \rho)$ pair per segment. Its kernels pin an NVIDIA Warp version whose successor changed an API they rely on, so we could not run the released solver unmodified and report no orbit sweep for it.

Visual vibration tomography (CVPR'22), released solver. Visual vibration tomography (Feng et al., 2022) recovers spatially-varying modulus *and* density fields from monocular video of surface

vibrations via $K(w)u = \omega^2 M(v)u$ with $K = \sum_e w_e K_e$, $M = \sum_e v_e M_e$—the linear-elastic special case of Proposition 2, which additionally covers the nonlinear plastic, granular and weakly-compressible blocks VVT does not treat. Its authors state the problem outright: "the objective function can become arbitrarily low if we do not constrain the range of material-property values; this is because scaling K and M by a factor of s still satisfies the generalized eigenvalue equation." They fix the gauge with an explicit penalty on the deviation of the stiffness field from a chosen mean value, and confine their claims to *relative* differences.

Running their released solver unmodified confirms both halves. Under the gauge the eigenvalues and the mode directions differ by *exactly* 0 (raw eigenvectors scale by $1/\sqrt{\alpha}$, an M-normalization convention with no observable content) and the eigen-equation residual scales by exactly α, so the feasible set is exactly invariant. And their estimator is exactly linear in the prior: scaling the prior means by 4 scales the recovered fields by 4.000 (maximum deviation 0) with shape correlation 1.000. This is structural rather than incidental—their `solve_w_v` forms $A = [A_{\text{constr}}; A_{\text{reg}}; A_{\text{mean}}]$ and solves $A^\top A\, x = A_{\text{mean}}^\top b_{\text{mean}}$, so the data enters only the normal matrix while the right-hand side is a pure function of the prior. Both released demos set the prior mean to the initialization.

VVT thus names its gauge choice and reports gauge-invariant (relative) quantities, whereas the systems of Table 1 that report absolute moduli share the same mass–stress structure with the scale fixed by an assumed density.

Resonance4D (concurrent; claims only, no code released). Resonance4D (Zhang et al., 2026) is closest to our analysis: it optimizes six part-level parameters—Young's modulus, Poisson's ratio, yield stress, plastic viscosity, internal friction angle *and density*—jointly, and is described as "preset-free." Table 3 partitions that set without ambiguity: ν and the friction angle are not removed by the mass-scale gauge; E, σ_Y, the plastic viscosity and ρ are four coordinates on a three-dimensional invariant quotient plus one gauge direction. Freeing ρ adds a coordinate along the gauge rather than information about it.

Two of their reported observations are consistent with the gauge. They note sensitivity to initialization, and that "different parameter combinations may induce substantially different motion patterns while yielding similarly plausible frame-wise appearances"—consistent with an orbit. They report no recovered absolute values. Their ablation finds that optimizing density improves the fit, which Proposition 2 suggests interpreting as follows: since motion along the orbit cannot change any appearance metric, the improvement must come from moving *across* the orbit—correcting wave speeds or the relative density between parts, which are observable—and not from measuring the absolute density. We flag this as a prediction rather than a finding, since no code is released; a ρ-versus-initialization scatter of the kind in §6 would settle it.

Measurement-grounded benchmarks and public datasets. We surveyed vision benchmarks and the experimental-modal-analysis literature for a dataset that could complete the validation, and found no public release that provides, simultaneously, a specimen-specific bench modulus, a trustworthy inversion regime, and passive video. Several provide two. GAUGE (Wang et al., 2026a) has the modulus—by mechanical testing with digital image correlation, plus measured densities and per-block masses—but its one deformable task with an apparent known load is actuator-driven at constant speed (53.2 ± 0.3 mm/s, median acceleration -0.00 m/s^2: prescribed displacement, not weight), and its only freely-moving deformable specimen is a foam/rubber composite settling through 37–70°, outside any small-deflection relation. IRIS (Khanbayov et al., 2026) has neither: its 240 real videos are rigid-body with kinematic and damping ground truth and no elastic constants. The experimental-mechanics releases invert the pattern—load and regime, no modulus. The closest is O'Donoughue et al. (2023): a uniform aluminium cantilever (264 mm free, 20×5 mm) struck by an instrumented hammer and filmed at 40 kfps, missing only a specimen-specific E. The gap is structural: Zaletelj et al. (2022) excite a free–free beam with a modal hammer and measure it at 919 high-speed-camera locations, yet fix ρ=7400 kg/m^3 and E=180 GPa at nominal values and update localized element parameters—the same anchor-fixing we audit, entirely defensible there, with no independent modulus to check against; and the openly released modal benchmarks we found are bolted frames (Dragonetti et al., 2025), where bolted joints and bracing, not material stiffness alone, set the response.

E REAL-VIDEO ROD STUDY

Data and estimator. Davis et al. (2017) released 2500 fps video of $1/8''$ metal rods clamped to a concrete block and excited by a loudspeaker, with a spreadsheet of per-rod properties: the densities were measured per rod and carry error bars, while the moduli are nominal manufacturer values without uncertainties. We stream every fifth frame of the 14.1 GB recordings over HTTP range requests, reduce each to a per-column vertical centroid, and recover c from the fundamental of a clamped–free rod, $f_1 = (\beta_1 L)^2 (8\pi)^{-1} (d/L^2)\, c$, where d is the rod diameter, L its free length and $\beta_1 L = 1.8751$ the first root of the clamped–free characteristic equation. The peak search is deliberately *prior-free*—the largest peak in a 2–30 Hz window set by the camera and the rod class, never by an assumed modulus—and a single detection gate (SNR ≥ 5, mode-shape correlation ≥ 0.4, not on a mains line) is applied to every rod at both clamped lengths.

A matched-wave-speed pair, already filmed. Two of the rods realise Prop. 1 physically. Aluminium (ρ=2696 kg/m^3) and steel (ρ=7773) differ by 2.88× in density and 2.90× in modulus, so their wave speeds agree to 0.3%: they sit on one orbit, and the proposition says they must vibrate alike. They do. The measured fundamentals, 6.858 and 6.837 Hz, lie 0.37 of one FFT bin apart—*indistinguishable* at our 0.057 Hz resolution—while brass and copper, off the orbit, resolve cleanly at 4.773 and 5.062 Hz (Fig. 3a). The coincidence belongs to the rods and not to the laboratory: neither brass nor copper shows any peak in that band (1.3× and 2.6× the median level, against 18× and 338× at their own fundamentals), which excludes a common environmental resonance. Nor is it camera shake: a lock-in at f_1 gives an amplitude climbing monotonically from clamp to tip with 14–32× dynamic range across the frame, correlating 0.58–0.86 with the analytic first bending mode (Fig. 3b), where whole-frame motion would be flat.

What the anchor buys, and where it stops. Conditioning each rod's c on its *measured* density gives absolute moduli tracking the reference across a 2.9× range: log–log slope 0.971, R^2=0.987 (n=4), median $|\log_{10}(\hat{E}/E_{\text{ref}})| = 0.027$ dex (6.3%) (Fig. 3c). That is the prescription working end to end on real video, and it is also where it stops. All four errors are negative (bias -0.025 dex), and propagating the stated density, diameter and length uncertainties covers the reference for three rods but not steel, which misses by -4.6σ. Steel is the rod whose inputs are known *best*—density to 1.6%, diameter to 0.8%—so its interval is narrowest and is first to expose what the anchor cannot absorb: a residual model error, most plausibly clamp compliance. A declared, uncertainty-carrying anchor is therefore *necessary and not sufficient*—it buys calibration only up to the fidelity of the forward model, and a better-known anchor makes the remaining error more visible, not less. The deficit is also not uniform (-0.2% for brass against -10.2% for steel), so it does not cancel in ratios in general: pairwise ratio errors run 1.2% to 10.9%.

A second geometry tests the residual. The same rods were also filmed clamped at $15''$. The matched pair sharpens—aluminium and steel now share a fundamental exactly (14.314 Hz, ratio 1.0000), with the density-anchored modulus ratio $+0.59\%$ at *both* lengths—and the slope is unchanged (0.974 against 0.971). Everything about the *absolute* scale gets worse, as a compliant root predicts: the negative bias more than doubles ($-0.025 \to -0.056$ dex) and the number of rods whose interval includes the reference falls from 3/4 to 2/3. A deficit $\propto EI/(kL)$ should grow by $22/15 = 1.47$ when the rod is shortened, and it grows by 1.61 (aluminium) and 1.48 (steel)—but by 17× for brass, whose $22''$ deficit of 0.19% makes the ratio ill-conditioned, and the same model fails to explain why the deficit does not scale with E across rods. We therefore report the residual as diagnosed, not characterised. Copper at $15''$ fails the detection gate—SNR 3.0 and shape correlation 0.20, against 45–276 and 0.72–0.96 elsewhere—and is excluded by the criterion that also excludes an unexcited higher mode.

Corroboration, and one trap. Our absolute errors reproduce those Davis et al. (2017) report for the same rods from the same footage (aluminium and steel: -9.1% and -10.2% against their -7.8% and -10.4%), reached with a different tracker and estimator. One artifact deserves naming: every recording carries a 119.99 Hz component with a 239.98 Hz harmonic—lighting flicker at four times the mains, identical across all rods to three decimals. For steel it falls within 0.01% of the predicted third bending mode and implies a wave speed agreeing with the true one to four digits, so an unguarded pipeline reports it as a cross-check. We reject it by frequency.

Validating an anchor by transfer. The rod study shows both what a declared anchor buys (slope 0.971 against independent references) and where it stops (a residual model error that a better-known anchor makes *more* visible). Completing it needs a specimen whose E is measured rather than quoted. The natural design is to predict that specimen's deflection under a held-out calibrated load. It does not work, and we say so because it is the design we first intended. For a uniform beam $f_1^2 \propto EI/(\rho AL^4)$ while the weighed mass is $m = \rho AL$, hence $EI \propto mL^3 f_1^2$: that beam's static bending stiffness follows from its own resonance and its own mass *without ever separating E from ρ*. A held-out force on the same geometry therefore tests the beam model and the two measurements that already produced the estimate; it never tests the attribution of a modulus to the *material*, because no step in the chain used a material-level fact beyond f_1 and m. The split has to be validated by *transfer*: recover c from one specimen, anchor with its measured density, and predict a *different* geometry, or a sister coupon cut from the same stock, where the geometric factor no longer cancels. The rest is standard practice—slender specimens ($L/h \gtrsim 100$) on free–free nodal supports, since clamp compliance is the fastest way to make a closed-form experiment secretly non-closed-form; resonance read under impulse excitation as in ASTM E1876 (ASTM International, 2022); thickness by micrometer, the bending stiffness being cubic in it. One pair makes the point at negligible cost: aluminium ($E \approx 69\,\text{GPa}$, $\rho \approx 2700$) and steel ($E \approx 200\,\text{GPa}$, $\rho \approx 7850$) have wave speeds within 0.2% of each other, so two specimens of one geometry vibrate at the same frequencies on camera while differing by $2.9\times$ in modulus—Prop. 1's orbit realised in two objects one can hold.

Limitations in full. (i) The classification assumes stress and plastic thresholds homogeneous of degree one in the stress-dimensioned parameters; rate-dependent and learned laws need separate treatment, though `gaugefix` applies unchanged since it tests candidate directions numerically rather than deriving them. (ii) Calibration is established *under the assumed simulator and noise model*. (iii) The rod study is real but partial: the reference values are manufacturer figures for the alloy, not measurements of *these* specimens, and the residual negative bias we attribute to clamp compliance is diagnosed rather than independently measured. We surveyed the vision benchmarks and the experimental-modal-analysis literature for a specimen with an independently measured E and a stated uncertainty and found no public release providing that together with a trustworthy inversion regime and passive video; Appendix D says how each candidate falls short. We therefore claim calibration of the prescribed procedure against independent references, not metrological validation against certified specimens. (iv) Our forward model is a homogeneous continuum and the real objects are not, which is why we report c as model-conditional.

F Proposed Physical Anchor-Switch Experiment (Not Yet Executed)

Every protocol comparison in §4 runs inside our own solver, which invites the fair objection that the simulator obeys a symmetry we programmed into it. This appendix proposes the same comparison on physical hardware: one steel rule, one camera, one balance, one micrometer, with the object, the imaging and the estimator held fixed and *only the protocol* changing. It is written before any data are taken, and we state plainly that **we have not executed it**: what follows is a design, an uncertainty budget and a set of falsifiable predictions, offered so that the protocol can be criticised or carried out — by us or by anyone — rather than as evidence already in hand. The analysis code is in the supplement, and Appendix F.1 validates the observation pipeline it depends on against rendered ground truth.

Design. A steel rule is clamped with $L = 300\,\text{mm}$ free ($b = 28\,\text{mm}$, $h = 1.0\,\text{mm}$), its own engraved markings supplying the in-frame metric calibration that anchors L. Predicted first bending frequency $f_0 = 9.06\,\text{Hz}$; with a weighed $M_1 = 50\,\text{g}$ at the tip, $f_M = 4.47\,\text{Hz}$ on the exact finite-tip-mass model, a mass ratio $M_1/m_{\text{beam}} = 0.758$.

A, unanchored. Deflect and release; record free decay. Predicted: a ridge along $(1, 1)$ in $(\log E, \log \rho)$.

C, base excitation. Move the *clamp* through a rich arbitrary hand trajectory, tracking two fiducials on it to recover translation and rotation, and the beam simultaneously. No actuator and no force measurement: the measured clamp trajectory *is* the Dirichlet datum, and in the moving frame the forcing is $-\rho a_{\text{base}}(t)$, density-proportional and hence in case (i). C supplies far

more kinematic information than A—seconds of measured broadband input and full output against one decay—and must still leave the gauge direction flat.

B, anchored. Attach M_1, its centre of mass *measured* rather than called "the tip".

D, static. Incremental tip deflection under the same weighed mass. A consistency test of the two beam models, the timebase and the clamp treatment—*not* an independent modulus, since $E_{\text{dyn}}/E_{\text{static}}$ cancels M, b, h^3 and L^3.

Blinding. The rule is not weighed before fitting. (E, ρ) are fitted from A and B only; the model and analysis are then frozen and written to disk; $m_{\text{pred}} = \rho b h L$ and the frequency of a second, never-fitted mass M_2 are predicted; and only then is the balance read.

Uncertainty. A 2×10^5-draw Monte Carlo through the exact estimator gives $\sigma(\log_{10} \rho) = 0.0040$ and $\sigma(\log_{10} E) = 0.0064$ dex with a micrometer, degrading to 0.022 and 0.066 dex with calipers. Thickness and width cancel exactly from the mass prediction ($\mathrm{d}\log m_{\text{pred}}/\mathrm{d}\log h = \mathrm{d}\log \rho/\mathrm{d}\log h + 1 = 0$), so the blinded prediction is sharper than E itself at 0.0037 dex.

Systematics matter more than the random budget, and a Monte Carlo cannot see them. Attaching the mass 3 mm inboard of the assumed point biases ρ by $+0.0128$ dex and an unweighed 0.5 g tracking marker by the same amount—each over three times the random σ. The Rayleigh 0.2427 effective-mass closed form, valid only for the unloaded mode shape, biases ρ by $+0.0030$ dex at this mass ratio, so inference uses the exact characteristic equation (Mahmoud, 2020) and a finite-element reference model rather than the closed form. A common frame-clock error cancels from ρ exactly—it is inferred from a frequency *ratio*—but enters E as f_0^2, which is what makes the static condition, carrying no timebase at all, a useful clock check. Clamp compliance acts as an unknown added free length: two lengths *calibrate* that two-parameter model whether or not it is true, so four lengths are used and one is predicted from the others.

Predictions fixed in advance. We predict (Fig. 21) that A's profile is flat along $(1, 1)$ in $(\log E, \log \rho)$ to $\Delta\chi^2 < 2.3$ (the 68% two-parameter threshold) over two decades under the stated model; that C, despite at least $10\times$ the independent observations and a $3\times$ improvement on c, shows no gauge-direction curvature beyond the pre-specified air and support model; that B reduces the gauge-direction uncertainty by at least $20\times$; that m_{pred} matches the balance within 2% and $f(M_2)$ falls inside its 95% predictive interval; and that a fit at one free length predicts the resonance at another without refitting. Curvature in C would indicate an accidental anchor—a marker, a clamp property, air, support coupling, a prior, or a bug—and not a false theorem: an equivariance result does not become wrong because a physical realisation violated one of its hypotheses.

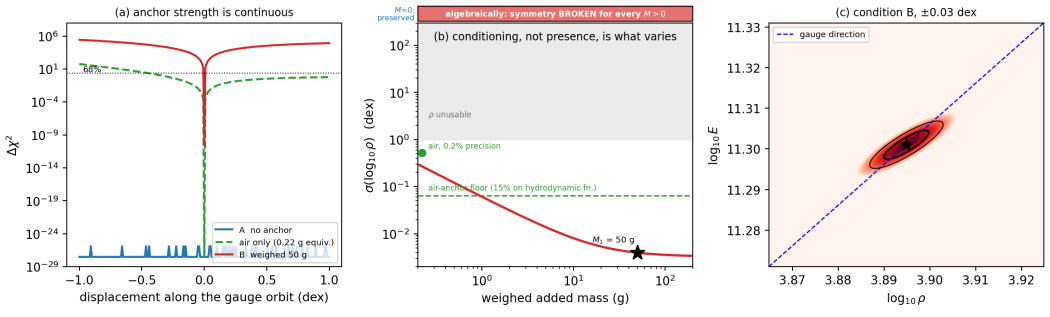

Figure 21: **Predicted anchor switch, on the exact beam model.** (a) Profile along the gauge orbit with no anchor (machine zero), with air alone, and with a weighed 50 g mass—thirty decades between the ends. (b) The algebraic answer is binary and identical for every $M > 0$ (strip), while the achievable precision on ρ varies over five decades of added mass; the air anchor sits far into the unusable region and saturates at the dashed floor set by the 15% uncertainty on the hydrodynamic function. (c) Condition B's confidence region, still elongated *along* the gauge direction: the anchor conditions that direction without making it the best-determined one.

F.1 PIPELINE VALIDATION ON SYNTHETIC VIDEO

The anchor switch of Appendix F rests on measuring two frequencies to 0.2% and on a systematics table whose entries are biases rather than variances. Neither can be checked against hardware, which supplies no ground truth. We therefore render the cantilever with an explicit camera model—analytic per-pixel coverage of the beam cross-section, integrated over the exposure window for motion blur and over row read-out time for rolling shutter, with photon and read noise quantised to 8 bits—and run the *same* pipeline used on the rod data (per-column intensity-weighted centroid, band-limited SVD modal coordinate, prior-free peak search, lock-in mode shape). Rendering our own forward model is closed-loop verification, so this validates the observation pipeline and says nothing about the protocol boundary; only hardware can do the latter.

It found a defect in our own estimator. Tracking is not the limiting factor—at 3.2 px beam thickness and 16 px tip amplitude the modal SNR is $\sim 1.7 \times 10^4$ and the recovered mode shape correlates at 1.000 with the analytic one. But parabolic interpolation of the Hann-windowed FFT peak, which the rod pipeline uses, carries a *bin-position-dependent* bias reaching -0.22%: it is $+0.047\%$ at 9.06 Hz and $+0.116\%$ at 4.47 Hz over a 10 s record, vanishing only when the frequency falls exactly on a bin. That exceeds the 0.2% the budget assumes and, worse, does *not* cancel from the ratio f_0/f_M that determines ρ, because the two conditions land at different bin offsets. Replacing it with a least-squares fit of a damped sinusoid removes it: exact to 10^{-7} on synthetic signals and -0.0097% through the full render-and-track chain, a $4.8\times$ improvement with a residual $20\times$ inside budget. The damping constant comes free and is useful for the air and clamp diagnostics. The rod results of Appendix E still use the parabolic estimator; for those records its bias is at most about half a percent in modulus, small against the aluminium and steel deficits of 9–10% but comparable to brass's -0.2%.

The remaining checks pass. A real `mp4v` encode–decode round trip changes the recovered frequency by 10^{-5} relative (-0.0097% raw versus -0.0098% encoded). Two rendered videos put through the full chain recover E to -0.00032 dex and ρ to -0.00024 dex, both an order of magnitude inside the random budget. The injected-bias predictions hold: a mass rendered 3.0 mm inboard of the assumed point and inverted as though at the tip gives $+0.0126$ dex on ρ against a predicted $+0.0128$, and a $+0.5\%$ frame-clock error leaves ρ at -0.0003 dex—the predicted exact cancellation, since ρ comes from a ratio—while moving E by -0.0048 dex against a predicted magnitude of 0.0043. Finally, orientation matters as predicted: with the beam along image columns, clamp and tip share sensor rows and the rolling-shutter phase gradient along the specimen is $+0.12°$, while along image rows it is $+7.99°$, $67\times$ larger. Single resonance estimates survive either way (-0.021% both), but condition C compares base and beam motion *within* a frame, so the beam is mounted along columns.

G REPRODUCIBILITY

All experiments use off-the-shelf differentiable renderers and simulators and run on a single A100 GPU. The NeuMA and Vid2Sim audits load the released checkpoints directly; the PhysGaussian audit loads its released ficus asset and simulates it with a standard MLS-MPM (we do not run PhysGaussian's own code). An aggregation script regenerates the NeuMA trajectory/render metrics from the saved rollouts. The real-video rod study is CPU-only and dominated by streaming about 23 GB of decimated high-speed footage (every fifth frame of eight 14.1 GB recordings).