PHYSON: PHYSICALIZING EVERYTHING EVERYWHERE ALL AT ONCE

Anonymous authors
Paper under double-blind review

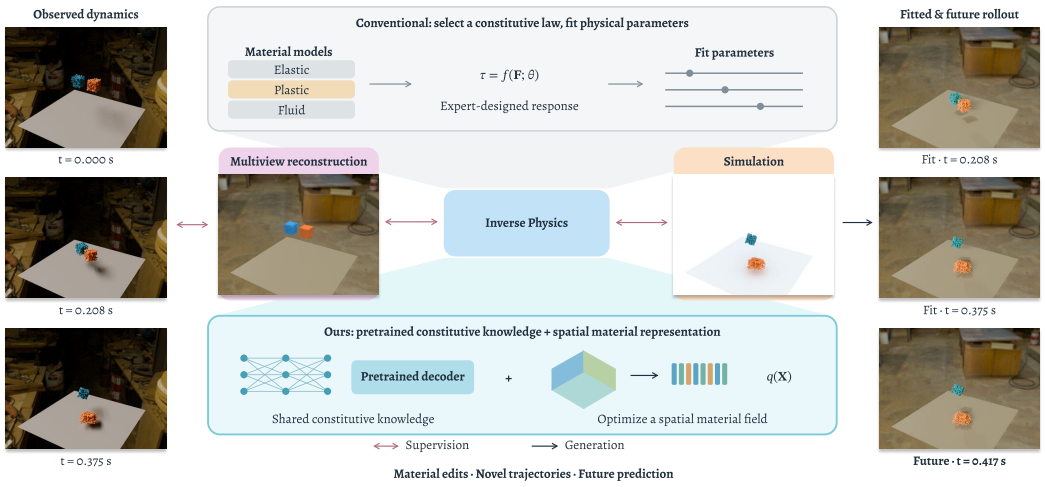


Figure 1: **PhysON** recovers spatially varying material behavior from multiview videos. Reconstructed geometry and particle trajectories guide the inference of a spatial material field decoded by a shared constitutive model, yielding simulation-ready assets for future prediction.

ABSTRACT

Recovering spatially varying material behavior from video requires identifying local responses through the coupled motion of a scene. We introduce PhysON, which combines a constitutive model pretrained across analytic material families with a per-particle material field. Reconstructed continuum trajectories and multiview silhouettes guide field optimization through differentiable simulation, while the shared constitutive model remains fixed. Observed-motion screening initializes the codes without region-wise material-family assignments. We also introduce PhysON-Validation, a synthetic multiview benchmark with paired on- and off-bank heterogeneous scenes. Position-supervised ablations separate the effects of initialization, gradient refinement, and spatial variation; constitutive-response diagnostics examine coverage beyond the training material instances. Together, these analyses distinguish what a shared prior can represent from what motion-based optimization can recover.

1 INTRODUCTION

Motion reveals how materials respond to forces. Recovering this behavior from video would turn captured scenes into physical assets for future prediction, robotic interaction, and physical scene editing (Genesis Authors, 2024; NVIDIA, 2025). Yet behavior can vary within an object and across interacting objects: different regions may deform, retain permanent changes, or dissipate motion differently. Identifying this variation requires more than assigning one material to the scene.

Visual system identification combines reconstruction and differentiable simulation (Li et al., 2023; Zhong et al., 2024; Cai et al., 2024). Neural constitutive models learn material responses (Ma et al., 2023; Cao et al., 2024; Mittal et al., 2025; Zhao et al., 2025), while spatial identification recovers

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.

nonuniform properties within specified models (Feng et al., 2022; Chen et al., 2022; Jiang et al., 2025; Yang et al., 2025b). OmniPhysGS (Lin et al., 2025) represents heterogeneous dynamics with weighted constitutive experts under video diffusion guidance. Our task is instead to infer spatially varying responses from captured motion using one shared learned constitutive representation.

The difficulty is that local material behavior is observed indirectly: each region's motion depends on both its response and forces transmitted by its surroundings. We address this coupled inverse problem with **PhysON** (Fig. 1). A constitutive model is pretrained across material families and then frozen. For a new scene, observed-motion screening selects a starting code, and differentiable MPM jointly refines one code per particle. The codes move with material points and represent variation within and across objects without region-wise material-family assignments.

We build on the continuum reconstruction of GIC and MASIV (Cai et al., 2024; Zhao et al., 2025): dynamic Gaussians provide geometry and motion, volume filling supplies simulation particles, and reconstructed trajectories and silhouettes supervise material inference. The recovered geometry and field are then reused for future prediction. We introduce PhysON-Validation to evaluate heterogeneous single objects and interacting objects, with paired material instances seen and unseen during pretraining.

Our contributions are:

- A per-particle material field decoded by a frozen shared constitutive prior, representing spatial variation across material families without region-wise family selection.
- A video-based inference framework that initializes and jointly fits this field from observed motion through differentiable simulation.
- PhysON-Validation, a synthetic multiview benchmark with trajectory and spatial material annotations and paired on- and off-bank scenes.

2 Related Work

Dynamic reconstruction. Dynamic NeRFs (Pumarola et al., 2021; Park et al., 2021; Fridovich-Keil et al., 2023) and Gaussian representations recover motion through deformation fields (Wu et al., 2024; Yang et al., 2024; Huang et al., 2024) or persistent primitives (Luiten et al., 2024). Learning-based methods estimate geometry and cameras (Wang et al., 2024; 2025a; Zhang et al., 2025; Wang et al., 2025c; Li et al., 2025; Wang et al., 2025b; Zhou et al., 2026; Zhang et al., 2026) or predict dynamic representations (Ren et al., 2024; Yang et al., 2025a). Material inference additionally needs physically scaled volume geometry and material-point trajectories. We use Gaussian-informed continuum reconstruction (Cai et al., 2024; Zhao et al., 2025) to obtain this supervision.

Physical understanding. NCLaw (Ma et al., 2023) learns constitutive laws through differentiable MPM. NeuMA (Cao et al., 2024) adapts a pretrained constitutive model with low-rank updates. UniPhy (Mittal et al., 2025) pretrains a shared latent-conditioned model and infers material codes with frozen weights. MASIV (Zhao et al., 2025) instead optimizes constitutive-network weights from video. We combine a frozen shared prior with a spatial field of particle codes.

Spatial identification is also established: Visual Vibration Tomography (Feng et al., 2022) recovers stiffness and density from surface vibrations, and Virtual Elastic Objects (Chen et al., 2022) fits inhomogeneous elasticity to multiview deformation. PhysTwin (Jiang et al., 2025) infers dense spring-mass properties, while PhysWorld (Yang et al., 2025b) selects constitutive models and refines per-particle physical parameters. Our field instead represents local responses through a common learned code space without selecting a family for each region.

NeuPhysics (Qiao et al., 2022), NeuroFluid (Guan et al., 2022), PAC-NeRF (Li et al., 2023), Spring-Gaus (Zhong et al., 2024), GIC (Cai et al., 2024), and Vid2Sim (Chen et al., 2025) connect visual reconstruction with physical inference. MASIV reconstructs dense continuum trajectories for constitutive fitting; MOSIV (Liu et al., 2026) extends continuum reconstruction to interacting objects with per-object Chamfer and silhouette objectives. We build on motion-based supervision to infer particle codes rather than per-scene network weights.

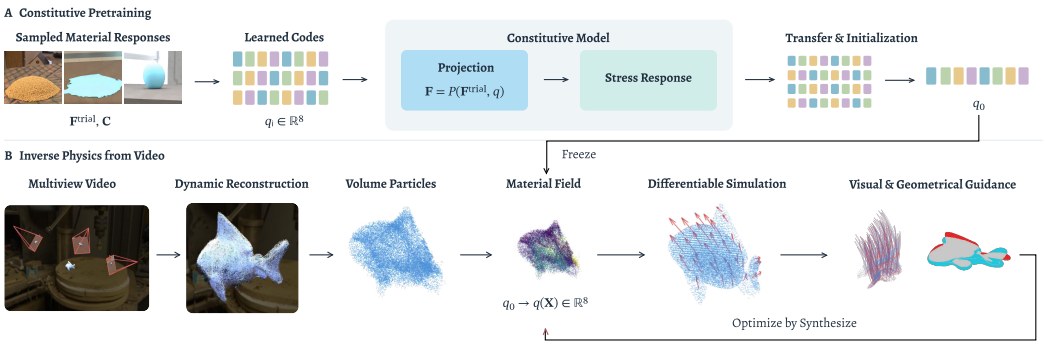


Figure 2: **Overview of PhysON.** Offline, analytic responses train a shared code-conditioned constitutive model. Online, reconstructed geometry and motion guide screening and per-particle code refinement through differentiable MPM. The constitutive heads remain frozen.

Physical generation. PhysGaussian (Xie et al., 2024) assigns material parameters to reconstructed Gaussians for forward simulation. PhysDreamer (Zhang et al., 2024), Physics3D (Liu et al., 2024a), and DreamPhysics (Huang et al., 2025) infer properties with video diffusion priors. OmniPhysGS (Lin et al., 2025) uses per-Gaussian mixtures of constitutive experts for heterogeneous dynamics generation. PhysGen (Liu et al., 2024b) and PhysMotion (Tan et al., 2026) use simulation to guide image-to-video generation. Our distinction is observed-motion inference of a spatial latent field decoded by a shared constitutive prior, rather than diffusion-guided generation.

3 Method

3.1 Overview

Given calibrated multiview videos $\{\mathbf{I}_t^v\}$ and foreground masks $\{\mathbf{M}_t^{v,k}\}_{k=1}^K$, we recover a spatial material field for a jointly simulated scene. Here $K = 1$ denotes a single object and $K > 1$ denotes interacting objects; either may contain heterogeneous materials. We first pretrain a shared constitutive representation, then freeze it and infer one code per reconstructed material point (Fig. 2).

Let $\bar{\mathbf{x}}(i)$ be the canonical position of material point i and $\mathbf{z}(i)$ its material code. Given the reconstructed initial state and scene boundary conditions $\mathcal{B}$, a differentiable MPM rollout produces particle states

$$\{\hat{\mathbf{x}}_n(i), \hat{\mathbf{v}}_n(i), \mathbf{F}_n^e(i), \mathbf{C}_n(i)\}_{i=1}^Q = \mathrm{MPM}\big(\bar{\mathbf{X}}, \mathbf{V}_0, \mathbf{z}; \mathcal{B}\big)_n , \tag{1}$$

where Q is the number of particles, n indexes simulation steps, and $\bar{\mathbf{X}}$ and $\mathbf{V}_0$ are the initial particle positions and velocities. Hats denote simulated positions and velocities; unadorned positions denote reconstructed trajectories. We optimize $\mathbf{z}$ so that the simulated motion agrees with the reconstructed 3D trajectories and their multiview silhouettes.

3.2 Learning a Unified Material Representation

Directly optimizing the weights of a constitutive network from a short video is severely underconstrained. Inspired by UniPhy (Mittal et al., 2025), we instead pretrain a material representation once and restrict per-scene inference to its code space. We sample material instances from analytic constitutive models covering five families: neo-Hookean elastic solids, von Mises plasticine, Drucker–Prager sand, Newtonian fluids, and viscoplastic non-Newtonian fluids. Each instance is simulated with MPM in several scenes (a fall, a shear, a compression, and a resting cube), and at every substep the analytic law records its inputs, the returned deformation gradient $\mathbf{F}^{e,*}$, and the Kirchhoff stress $\boldsymbol{\tau}^*$. The states are therefore those that each material actually visits under large deformation, contact, and flow.

We represent the constitutive update with a projection head $\mathcal{P}_{\boldsymbol{\theta}^p}$ and a stress head $\mathcal{S}_{\boldsymbol{\theta}^e}$, both conditioned on the same learned material code $\mathbf{z} \in \mathbb{R}^d$. Within one substep, the stress is evaluated on the current state, MPM transfers advance positions, velocities, and the trial deformation gradient, and

the projection returns the next elastic deformation gradient:

$$\boldsymbol{\tau}_n(i) = \mathcal{S}_{\boldsymbol{\theta}^e}(\mathbf{F}^e_n(i), \mathbf{C}_n(i); \mathbf{z}(i)), \tag{2}$$

$$\mathbf{F}^e_{n+1}(i) = \mathcal{P}_{\boldsymbol{\theta}^p}\left(\mathbf{F}^{e,\text{trial}}_{n+1}(i); \mathbf{z}(i)\right), \qquad \mathbf{F}^{e,\text{trial}}_{n+1}(i) = (\mathbf{I} + \Delta t\, \mathbf{C}_{n+1}(i))\, \mathbf{F}^e_n(i). \tag{3}$$

Here $\mathbf{F}^e$ is the elastic deformation gradient, Δt is the substep duration, and $\mathbf{C}$ is the affine velocity gradient. Both heads use the singular value decomposition $\mathbf{F} = \mathbf{U}\,\mathrm{diag}(\boldsymbol{\sigma})\mathbf{V}^\top$ with singular values clamped below at 0.05, and the rotation-invariant features $\boldsymbol{\sigma} - \mathbf{1}$, $\mathbf{F}^\top\mathbf{F} - \mathbf{I}$, and $\det\mathbf{F} - 1$; the stress head additionally receives the symmetric part of $\mathbf{C}$ in the rotated frame, $\mathbf{R}^\top\,\mathrm{sym}(\mathbf{C})\mathbf{R}$ with $\mathbf{R} = \mathbf{U}\mathbf{V}^\top$. Each head is a multilayer perceptron on these features concatenated with $\mathbf{z}$ that outputs a symmetric matrix $\mathbf{S}$. The projection returns $\mathbf{F}^e = \mathbf{U}\,\mathrm{diag}(\boldsymbol{\sigma})\mathbf{V}^\top + 10^{-3}\mathbf{R}\mathbf{S}_p$, a small learned correction to the clamped input, and the stress is $\boldsymbol{\tau} = \mathbf{R}\mathbf{S}_e\mathbf{F}^\top$. The heads use the current state only, without a learned history variable.

We jointly learn the head parameters and one code $\mathbf{z}_m$ per training material instance m with the constitutive objective

$$\mathcal{L}_{\text{const}} = \mathcal{L}_{\text{stress}} + \mathcal{L}_{\text{proj}} + \lambda_z\mathcal{L}_z + \lambda_\Sigma\mathcal{L}_\Sigma, \tag{4}$$

where $\mathcal{L}_{\text{stress}}$ and $\mathcal{L}_{\text{proj}}$ are normalized squared errors of the stress and of the projection correction $\mathbf{F}^e - \mathbf{F}^{e,\text{trial}}$ against the analytic targets, $\mathcal{L}_z$ is the mean code norm, and $\mathcal{L}_\Sigma$ keeps the covariance of the code bank near a fixed isotropic scale; definitions and weights are given in Appendix B. During pretraining, the stress head receives the analytic returned state $(\mathbf{F}^{e,*}, \mathbf{C})$ as input, so the two heads are supervised independently. After pretraining, $\boldsymbol{\theta}^p$ and $\boldsymbol{\theta}^e$ are frozen; for a new scene only material codes are optimized. Each particle is therefore optimized in a compact code space rather than through its own constitutive network. We evaluate response coverage and inference on held-out material parameters in Sec. 5.3.

3.3 From Dynamic Gaussians to a Continuum

We first fit a dynamic Gaussian representation (Wu et al., 2024) to the input videos. The reconstruction provides canonical Gaussians for the scene and a time-dependent deformation that explains their motion across frames. This intermediate representation supplies multiview-consistent geometry, appearance, and persistent correspondences, but it is concentrated near the visible surfaces and therefore cannot be used directly by a continuum simulator.

We follow the volume preparation of MASIV (Zhao et al., 2025) and GIC (Cai et al., 2024): multiview masks and depths define an occupancy volume, which is smoothed and thresholded to obtain material points. Where material volume is declared, it guides the threshold choice. All objects' particles are simulated jointly. The filling procedure and its sensitivity are detailed in Appendices B.5 and D.2.

We fit canonical particle positions and a deformation network to the per-frame filled volumes using bidirectional Chamfer distance. Advecting these particles gives target positions $\mathbf{X}_t$ with persistent identity. Before material inference, we fit a shared initial-velocity vector to the first four observed frames, or use the declared initial velocity field on PhysON-Validation. Reconstruction and initial velocity remain fixed during material fitting; schedules are given in Appendix B.6.

3.4 Per-Particle Material Codes

A single code per scene cannot represent variation within an object or different responses among several objects. We instead give every material point its own code $\mathbf{z}(i) \in \mathbb{R}^d$, attached to the particle's identity rather than to its current position, so that material is advected with each object. The codes $\mathbf{Z} = [\mathbf{z}(1), \ldots, \mathbf{z}(Q)]^\top$ are the only unknowns of system identification. They represent variation both within and across objects without object identities or material labels as inputs. The parameterization imposes no spatial smoothness; neighboring particles exchange momentum through the shared grid, coupling their effects on the observation loss. A homogeneous scene is the special case in which all codes are equal.

At every MPM substep, Eqs. 2–3 are queried with the particle-specific code $\mathbf{z}(i)$. Standard particle-to-grid transfer, grid update, boundary handling, and grid-to-particle transfer then advance the state. Thus, gradients from a mismatch at any later frame propagate through the full rollout to the code of each particle.

Table 1: **PhysON-Validation scenes.** Each scene exists in an on-bank and an off-bank variant with identical geometry, phase layout and initial state (14 trajectories). Speed is the initial particle speed; in the water scene the pool starts at rest.

Scene	Objects	Particles	Phases	Material	Varied parameters	Speed (m/s)
Elastic cube	1	1,000	5	elastic	μ, λ	0.32
Plasticine block	1	2,000	5	plasticine	μ, λ, yield stress	0.47
Sand column	1	2,400	5	sand	μ, λ, friction	0.43
Newtonian block	1	10,000	3	Newtonian	viscosity, bulk modulus	0.20
Non-Newtonian sphere	1	5,185	3	non-Newtonian	μ, yield stress, viscosity	0.53
Elastic–plastic collision	2	500 + 500	2	elastic, plasticine	μ, λ, yield stress	0.54
Elastic cube into water	2	2,744 + 46,575	2	elastic, Newtonian	μ, λ, viscosity	1.5 / 0

3.5 Geometry-Aligned System Identification

For every scene, the dynamic reconstruction supplies volumetric particle trajectories with persistent correspondences. We align simulated particles with their reconstructed counterparts using a trajectory L1 loss over the observed window. This correspondence-aware objective captures particle motion, including tangential sliding, which set-based distances such as Chamfer can miss because they do not enforce point-wise correspondence (Zhao et al., 2025):

$$\mathcal{L}_{\text{geo}} = \mathcal{L}_1\Big(\{\hat{\mathbf{X}}_n\}_{n=1}^{N}, \{\mathbf{X}_n\}_{n=1}^{N}\Big), \tag{5}$$

where $\mathcal{L}_1$ denotes mean absolute error, N is the number of observed simulation substeps, and $\hat{\mathbf{X}}_n$ and $\mathbf{X}_n$ contain the simulated and reconstructed particle positions in matching order, the targets being interpolated to every substep. Positions in this loss are expressed in the reconstruction coordinate system. Chamfer distance is retained as an evaluation metric. Following MASIV (Zhao et al., 2025), we also constrain the simulated silhouettes:

$$\mathcal{L}_{\text{sil}} = \mathcal{L}_1\Big(\{\hat{\mathbf{M}}_t^v\}_{(t,v)\in\mathcal{O}}, \{\mathbf{M}_t^v\}_{(t,v)\in\mathcal{O}}\Big). \tag{6}$$

Here $\mathcal{O}$ contains the observed frame–view pairs, $\hat{\mathbf{M}}_t^v = \mathcal{R}_\alpha(\hat{\mathbf{X}}_t; v)$ is the alpha map rendered from one isotropic Gaussian per simulated particle, and $\mathbf{M}_t^v$ is the foreground mask of the input image; the loss covers the full image. The geometric loss constrains particle motion, while the silhouette loss constrains the projected scene boundary.

The complete per-scene objective is

$$\min_{\mathbf{Z}} \quad \lambda_{\text{geo}}\mathcal{L}_{\text{geo}} + \lambda_{\text{sil}}\mathcal{L}_{\text{sil}}, \tag{7}$$

with $\lambda_{\text{geo}} = 10$ and $\lambda_{\text{sil}} = 1$ in all experiments. Initialization matters in practice (Table 6). We initialize by screening. We assign each of $K = 32$ codes, evenly spaced over the pretrained bank, to every particle, run one observed-window rollout without gradients, and start from the code with the lowest objective. Screening uses only observed frames and no material labels. We then optimize all particle codes with Adam through differentiable MPM, keeping the constitutive heads frozen. An update that produces a nonfinite objective or gradient is rejected and retried at half the learning rate (Appendix B.2). The output is the code field with the lowest observed objective; future frames are used only for evaluation. The selected field and geometry define the simulation-ready scene.

4 Dataset

PhysON-Validation contains 14 synthetic multiview trajectories: seven scenes, each with an on-bank and an off-bank variant. Five scenes contain a single heterogeneous object with three to five material phases: an elastic cube, plasticine block, sand column, Newtonian block, and non-Newtonian sphere. Two scenes contain interacting objects: elastic–plastic collision and an elastic cube entering water. Table 1 lists their geometry, particle counts, and material variation.

On-bank phases use exact parameters of the 200 pretraining material instances. Off-bank counterparts preserve geometry, phase layout, initial state, and categorical laws but change numeric parameters to values absent from the bank. Thus the comparison tests unseen material parameters within the pretrained families, with spatial layout and scene conditions controlled.

		PAC-NeRF						Spring-Gaus							
	Method	Newtonian	Non-Newtonian	Elasticity	Plasticine	Sand	Mean	Torus	Cross	Cream	Apple	Paste	Chess	Banana	Mean
CD ↓	PAC-NeRF Li et al. (2023)	0.277	0.236	0.238	0.429	**0.212**	0.278	4.92	1.10	0.77	1.11	3.14	0.96	2.77	2.11
	Spring-Gaus Zhong et al. (2024)	-	-	-	-	-	-	0.17	0.48	0.36	0.38	0.19	1.80	2.60	0.85
	NeuMA Cao et al. (2024)	-	-	-	-	-	-	4.59	**0.06**	**0.08**	**0.04**	0.71	**0.05**	**0.03**	0.79
	GIC Cai et al. (2024)	0.243	**0.195**	**0.178**	**0.196**	0.250	0.212	0.13	0.13	0.14	0.15	0.17	0.41	**0.03**	0.17
	MASIV Zhao et al. (2025)	**0.233**	0.198	0.192	0.201	0.229	**0.210**	**0.08**	0.10	0.19	0.16	**0.13**	0.19	0.08	**0.13**
	PhysON (Ours)	0.083	0.071	0.020	0.057	0.109	0.068	0.03	0.04	0.07	0.22	0.03	0.29	0.18	0.12

Table 2: **Observable state simulation on PAC-NeRF and Spring-Gaus Synthetic datasets.** The current PhysON row measures distance to its reconstructed fitting targets, not the ground-truth particles used for the baselines; its values are not directly comparable.

Each trajectory has 30 frames rendered from 11 calibrated cameras at 800×800. The first 20 frames are observed and the final 10 are held out. Archives include particle identities, trajectories, material phases, and constitutive states for evaluation and analysis. Video-based fitting uses images, masks, cameras, timestamps, declared initial velocity, and scene settings including material volume, not the ground-truth material annotations. Simulation, rendering, phase parameters, and data-generation details are in Appendix C.

5 Experiments

5.1 Experimental Settings

Benchmarks and baselines. We evaluate on PhysON-Validation (14 paired on-/off-bank trajectories), PAC-NeRF (Li et al., 2023) (45 sequences across five material classes), and Spring-Gaus (Zhong et al., 2024) (7 synthetic scenes and 5 real captures). Public-benchmark baselines are PAC-NeRF, Spring-Gaus, NeuMA (Cao et al., 2024), GIC (Cai et al., 2024), and MASIV (Zhao et al., 2025). They use prescribed constitutive models, spring-mass representations, or adapted neural constitutive models; our method fits particle codes under a frozen shared prior. PhysON-Validation tables include MOSIV (Liu et al., 2026) and OmniPhysGS (Lin et al., 2025), with uncompleted entries marked [xx].

Metrics and protocol. We report observed and future errors separately: Chamfer Distance (CD, $10^3\,\mathrm{mm}^2$) measures geometry, while PSNR and SSIM measure rendering. PhysON-Validation uses all 11 views; public-benchmark image metrics follow the evaluation view recorded with their results. Real captures have no ground-truth particles. The current public observed-CD row for our method uses reconstructed targets rather than ground-truth particles, as noted in Table 2; it cannot establish a geometric improvement over the baselines.

Both constitutive heads use five hidden layers of width 256 and an eight-dimensional material code, pretrained on 200 material instances and 800 trajectories for 1M updates. Video-based fitting selects the field by observed objective and evaluates the future without resetting the state. The main screened protocol uses 300 material updates on public benchmarks and 60 on PhysON-Validation (40 for the water pool). Complete schedules, learning rates, reconstruction and appearance fitting are in Appendix B.6. The currently populated public image results use the earlier random-start protocol; they are not results of screened initialization.

5.2 Comparison Results

Public benchmarks. Table 3 reports future prediction on Spring-Gaus Synthetic. The current random-start results give mean CD 1.73, PSNR 16.94 and SSIM 0.902; CD ranges from 0.29 on torus to 5.29 on banana, and is the lowest among the compared methods on cream and toothpaste. Observed image results on the five real captures are reported in Appendix C.5 (Table 11). These results must be distinguished from the screened protocol evaluated below.

PhysON-Validation. Tables 4 and 5 separate on-bank scenes from off-bank counterparts with unchanged geometry and phase layout. The latter test parameters absent from pretraining, not new constitutive families. Entries marked ‡, and means that include them, are the best results so far for fits that are being rerun; pending entries are still running. Complete scene-level reporting and qualitative trajectories are provided in Appendix D.

	Method	Torus	Cross	Cream	Apple	Paste	Chess	Banana	Mean
CD ↓	PAC-NeRF Li et al. (2023)	2.47	3.87	2.21	4.69	37.70	8.20	66.43	17.94
	Spring-Gaus Zhong et al. (2024)	2.38	1.57	2.22	1.87	7.03	2.59	18.48	5.16
	NeuMA Cao et al. (2024)	1.17	0.40	1147.4	**0.05**	44.26	**0.33**	0.30	170.56
	GIC Cai et al. (2024)	0.75	1.09	0.94	0.22	2.79	0.77	**0.12**	**0.95**
	MASIV Zhao et al. (2025)	**0.16**	**0.38**	1.51	2.64	1.75	1.70	0.24	1.20
	PhysON (ours)	0.29	1.78	**0.53**	0.32	**1.62**	2.31	5.29	1.73
PSNR ↑	PAC-NeRF Li et al. (2023)	17.46	14.15	15.37	19.94	12.32	15.08	16.04	15.77
	Spring-Gaus Zhong et al. (2024)	16.83	16.93	15.42	21.55	14.71	16.08	17.89	17.06
	NeuMA Cao et al. (2024)	16.61	19.67	2.32	26.25	13.42	**20.85**	23.16	17.47
	GIC Cai et al. (2024)	20.24	30.51	**19.15**	**26.89**	**16.31**	18.44	**29.29**	**22.98**
	MASIV Zhao et al. (2025)	**22.42**	**33.73**	16.45	21.19	15.73	17.28	26.56	21.91
	PhysON (ours)	16.67	15.31	16.15	21.56	14.74	14.61	19.53	16.94
SSIM ↑	PAC-NeRF Li et al. (2023)	0.919	0.906	0.858	0.878	0.819	0.848	0.886	0.870
	Spring-Gaus Zhong et al. (2024)	0.919	0.940	0.862	0.902	0.872	0.881	0.904	0.897
	NeuMA Cao et al. (2024)	0.942	0.948	0.889	**0.964**	0.889	**0.933**	0.964	**0.933**
	GIC Cai et al. (2024)	0.942	0.939	**0.909**	0.948	**0.894**	0.912	0.964	0.930
	MASIV Zhao et al. (2025)	**0.955**	**0.967**	0.852	0.898	0.891	0.878	**0.965**	0.915
	PhysON (ours)	0.915	0.932	0.903	0.925	0.876	0.838	0.926	0.902

Table 3: **Future state simulation on Spring-Gaus Synthetic dataset.** Each rollout is fitted on the first 20 of 30 frames and continued from the initial state without reset through the last 10, which are never used for fitting or selection; values average these held-out frames. CD in $10^3\,\mathrm{mm}^2$; PSNR and SSIM on the MASIV evaluation view. Baseline results are from MASIV (Zhao et al., 2025). Paste: toothpaste. The PhysON rows use an earlier configuration without screening: codes start from the bank statistics and are optimized for 300 updates at learning rate 10^{-3}.

		On-bank								Off-bank							
	Method	Elastic	Plasticine	Sand	Newt.	Non-Newt.	El.–Pl.	El.–Water	Mean	Elastic	Plasticine	Sand	Newt.	Non-Newt.	El.–Pl.	El.–Water	Mean
CD ↓	MOSIV Liu et al. (2026)	1.193	1.630	106.5	76.82	88.95	241.2	25.76	77.43	[xx]	0.471	80.12	[xx]	132.7	1.861	[xx]	[xx]
	OmniPhysGS Lin et al. (2025)	2.563	0.873	0.523	3.857	13.50	30.08	3.347	7.821	4.427⋆	5.078	35.64	1.621⋆	21.56	25.35	[xx]	[xx]
	PhysON (ours)	0.147	0.122	0.305‡	1.043	0.201	0.130	0.306	0.322‡	0.135	0.122	1.142‡	0.234‡	0.185	0.808‡	0.260	0.412‡
PSNR ↑	MOSIV Liu et al. (2026)	22.66	20.00	13.18	9.39	11.91	16.35	8.61	14.59	[xx]	21.58	13.09	[xx]	12.24	18.79	[xx]	[xx]
	OmniPhysGS Lin et al. (2025)	22.61	22.26	19.94	14.69	20.19	16.11	15.75	18.79	26.37⋆	22.82	18.97	21.96⋆	18.60	18.76	[xx]	[xx]
	PhysON (ours)	37.06	33.52	25.80‡	23.40	25.92	34.70	20.71	28.73‡	36.67	31.65	24.36‡	28.12‡	28.75	31.54‡	26.00	29.58‡
SSIM ↑	MOSIV Liu et al. (2026)	0.984	0.974	0.818	0.786	0.860	0.962	0.737	0.874	[xx]	0.974	0.791	[xx]	0.879	0.969	[xx]	[xx]
	OmniPhysGS Lin et al. (2025)	0.982	0.973	0.913	0.836	0.928	0.897	0.879	0.915	0.974⋆	0.974	0.888	0.892⋆	0.796	0.966	[xx]	[xx]
	PhysON (ours)	0.994	0.990	0.950‡	0.914	0.957	0.990	0.940	0.962‡	0.994	0.988	0.914‡	0.970‡	0.975	0.985‡	0.959	0.969‡

Table 4: **Observable state simulation on PhysON-Validation.** Frames 1–19 of each 30-frame sequence, the frames used for fitting; frame 0, the reconstructed initial state, is excluded. On-bank scenes use pretraining material parameters; off-bank scenes share their geometry, phase layout and initial state but use parameters absent from the pretraining bank. El.–Pl.: elastic and plastic objects colliding; El.–Water: elastic object in a water pool. CD: mean of the two directed mean squared nearest-neighbor distances between all ground-truth and all simulated particles, in $10^3\,\mathrm{mm}^2$, averaged over frames; PSNR and SSIM average all 11 views for PhysON; MOSIV and OmniPhysGS render only the held-out camera, on which their PSNR and SSIM are computed. Mean: over the seven scenes of a variant. PhysON: per-particle codes start from the best of 32 bank codes by observed rollout and are optimized for 60 updates at learning rate 10^{-2} (3×10^{-3} for the non-Newtonian scene; 40 updates at 3×10^{-3} for the water pool); the appearance is fitted to the observed images and rendered on the simulated particles. ⋆OmniPhysGS run initialized from our reconstruction, evaluated at an intermediate checkpoint (60 or 90 of 120 updates) with PSNR and SSIM over all 11 views; used where no other OmniPhysGS result exists. ‡Provisional: best result so far for fits that are being rerun; these values and the means that include them will be updated.

5.3 Ablation Study

Initialization and refinement. Table 6 isolates material fitting from reconstruction error by supervising observed particle positions on the heterogeneous elastic scene. At the same learning rate of 10^{-3}, screening followed by per-particle optimization gives future L1 of 2.44 mm on-bank, compared with 59.41 mm from the random start. Screening alone gives 8.92 mm: refinement remains necessary. The screened run at 10^{-2} reaches 1.68 mm. Off-bank, the corresponding result is 2.17 mm versus 9.86 mm from the random start, while the screened code without refinement gives

		On-bank								Off-bank							
	Method	Elastic	Plasticine	Sand	Newt.	Non-Newt.	El.–Pl.	El.–Water	Mean	Elastic	Plasticine	Sand	Newt.	Non-Newt.	El.–Pl.	El.–Water	Mean
CD ↓	MOSIV Liu et al. (2026)	18.82	13.85	1349	854.0	1132	288.0	25.14	525.9	[xx]	5.422	1175	[xx]	1472	0.908	[xx]	[xx]
	OmniPhysGS Lin et al. (2025)	44.23	1.596	3.266	6.662	25.96	57.44	5.802	20.71	3.310*	8.437	47.00	2.138*	58.37	69.41	[xx]	[xx]
	PhysON (ours)	1.832	0.207	0.303‡	1.057	0.153	0.220	0.941	0.673‡	0.226	0.142	1.799‡	0.206‡	0.176	0.258‡	0.435	0.463‡
PSNR ↑	MOSIV Liu et al. (2026)	19.05	17.23	9.70	6.00	11.48	16.61	8.09	12.59	[xx]	18.80	9.44	[xx]	12.24	19.01	[xx]	[xx]
	OmniPhysGS Lin et al. (2025)	18.59	18.79	14.01	13.69	13.29	12.43	13.88	14.96	22.08*	19.57	12.85	16.93*	12.44	16.00	[xx]	[xx]
	PhysON (ours)	26.49	32.77	22.84‡	21.63	25.94	30.60	16.15	25.20‡	30.51	31.67	20.60‡	25.96‡	28.53	31.15‡	22.32	27.25‡
SSIM ↑	MOSIV Liu et al. (2026)	0.978	0.965	0.536	0.665	0.861	0.964	0.724	0.813	[xx]	0.964	0.503	[xx]	0.881	0.971	[xx]	[xx]
	OmniPhysGS Lin et al. (2025)	0.973	0.958	0.863	0.847	0.893	0.766	0.867	0.881	0.970*	0.965	0.820	0.854*	0.741	0.956	[xx]	[xx]
	PhysON (ours)	0.980	0.988	0.930‡	0.907	0.946	0.984	0.908	0.949‡	0.989	0.987	0.868‡	0.962‡	0.972	0.987‡	0.942	0.958‡

Table 5: **Future state simulation on PhysON-Validation.** Held-out frames 20–29, predicted by continuing each fitted rollout from the initial state without any reset. Future frames are never used for fitting, selection or appearance fitting. Scenes, metrics and settings as in Table 4.

Table 6: **Material-code initialization and optimization** on PhysON-Validation elastic (five stiffness layers, 1,000 particles), fitted to ground-truth particle positions of observed frames 1–19 and evaluated on held-out frames 20–29. The on- and off-bank scenes share geometry, layers and initial state; off-bank layer parameters are absent from the pretraining bank. Frozen Q8 head, one code per particle unless stated, 300 updates, observed-only selection. The indented random-start variants were run on the on-bank scene only. L1 in mm, CD in $10^3\,\mathrm{mm}^2$.

	On-bank				Off-bank			
Variant	Obs. L1	Fut. L1	Fut. CD	Final L1	Obs. L1	Fut. L1	Fut. CD	Final L1
Random start (bank statistics), lr 10^{-3}	15.81	59.41	4.298	81.98	2.71	9.86	0.132	12.34
isotropic random start	16.00	59.45	4.263	81.67	–	–	–	–
lr 10^{-2}	13.39	50.03	2.738	72.67	–	–	–	–
response-coordinate Adam	16.50	60.40	4.192	79.26	–	–	–	–
+ head fine-tuning (lr 10^{-5})	9.93	22.29	0.916	36.14	–	–	–	–
Screened code only (0 updates)	3.47	8.92	0.114	4.26	7.13	28.12	0.925	12.94
Screened start, one shared code	1.49	15.54	0.245	14.42	1.53†	4.86†	0.057†	9.72†
Screened start, lr 10^{-3}	0.19	2.44	0.039	3.43	0.19	2.18	0.026	2.96
Screened start, lr 10^{-2} (PhysON)	0.21	**1.68**	**0.016**	**2.80**	0.24	**2.17**	**0.020**	**2.45**
Known bank code (ground-truth assisted)	0.75	3.41	0.060	2.34	–	–	–	–

†Stopped at update 251 by the optimizer plateau rule. – : not run, or no bank code exists.

Table 7: **Pretrained constitutive heads on materials outside the bank.** Stress relative error on the held-out (future-frame) states of each PhysON-Validation single-object scene, averaged over material phases. On-bank: the exact bank code. Off-bank: the bank code nearest in log parameters, or one code per phase fitted to the teacher responses of the observed-frame states (Adam, 1,500 updates, selected on those states). Uses ground-truth responses; a Stage 1 analysis, never a Stage 2 input.

	Bank validation	On-bank	Off-bank		Off-bank fitted
Head	loss ($\times 10^{-5}$)	exact code	nearest code	fitted code	total loss ($\times 10^{-5}$)
Q8 (codes into the heads)	9.02	0.113	0.296	0.106	5.2
Q8L64 (codes lifted to 64-D)	10.77	0.119	0.297	0.105	36.8

28.12 mm. Larger learning rates, isotropic initialization, and response-coordinate Adam do not close the on-bank random-start gap; head fine-tuning gives 22.29 mm.

Spatial representation and constitutive coverage. From a screened start, the on-bank shared-code control gives 15.54 mm future L1, compared with 2.44 mm for the per-particle field at 10^{-3}. The representation therefore matters beyond selecting a bank entry. The ground-truth-assisted bank-code result is 3.41 mm and is a diagnostic, not an inference baseline. Table 7 reports the Q8/Q8L64 response study: off-bank codes fitted to teacher responses achieve 0.106 relative stress error on held-out states versus 0.296 for the nearest bank code. This is evidence of response coverage on those states, distinct from position-based recovery. Analytic-head replacements diagnose the stress and projection components separately.

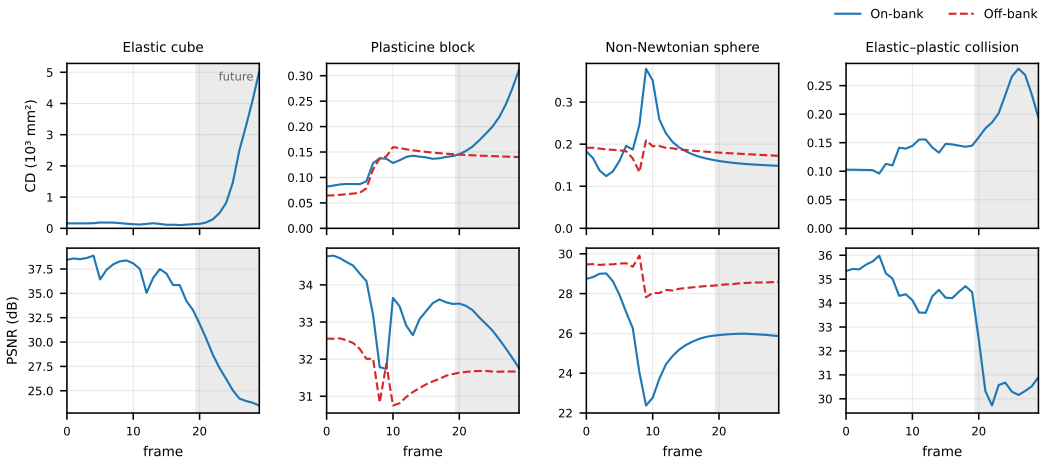


Figure 3: **Per-frame errors on PhysON-Validation.** Top: Chamfer distance ($10^3\,\mathrm{mm}^2$) between all ground-truth and all simulated particles. Bottom: PSNR (dB), averaged over the 11 views, of the appearance fitted on observed images and rendered on the simulated particles. Solid: on-bank; dashed: off-bank. The gray band marks the held-out future frames 20–29. At frame 0 the Chamfer distance is the error of the reconstructed initial geometry. Missing curves (sand, Newtonian, water pool; off-bank elastic cube and collision) and the MOSIV and OmniPhysGS curves: [xx].

5.4 Qualitative Analysis

Appendix D.4 shows observed and future RGB and particle trajectories. The plasticine examples reproduce impact and spreading, whereas the elastic cube and collision expose errors in later rotation and separation. These comparisons distinguish a close observed fit from a reliable continuation. New-initial-velocity tests and their limitations are reported in Table 14 and Fig. 8; future continuation alone is not evidence of generalization to a new interaction.

Figure 8 replays three video-fitted fields under changed initial velocities without refitting. Plasticine transfer is partial: on-bank mean CD over frames 1–29 is 0.276 for the downward change and 0.154 for the lateral change, versus 0.151 originally. Off-bank plasticine gives 0.133 and 0.220 for the same changes, compared with 0.129 at the original velocity. Transfer fails on the elastic–plastic collision: its downward rollout becomes nonfinite at frame 23, while the lateral change gives mean CD 0.822 versus 0.161 originally, with simulated slabs staying in contact instead of bouncing apart. Appendix D.5 gives the full protocol and particle comparisons.

6 Conclusion

We presented PhysON, which combines a frozen constitutive prior pretrained across material families with a spatial field of particle codes inferred from multiview video. Observed-motion screening provides a starting point for differentiable refinement without assigning material families to regions. PhysON-Validation adds paired on- and off-bank heterogeneous scenes. Our analyses distinguish representing material responses from recovering them through motion and identify initialization and spatial degrees of freedom as central to inference.

Limitation Discussion. The inferred codes can absorb reconstruction and initial-state errors, and weakly excited behavior remains ambiguous. The prior is limited by its analytic families and state assumptions; screening does not ensure convergence or reliable prediction under new interactions. Contact, long horizons, and undeformed appearance kernels remain challenging. PhysON-Validation shares the simulator and analytic laws used in pretraining, and real evaluation is limited to observed image metrics on five captures. Full-rollout differentiation is expensive.

ETHICS STATEMENT

This work studies material inference using simulated scenes and public visual system-identification benchmarks. Its intended applications include physical scene reconstruction and simulation for graphics and robotics. Inferred dynamics are approximate and should not be treated as certified material measurements or deployed in safety-critical control without independent validation and appropriate safeguards. Physically plausible rendered motion could also be mistaken for evidence of real events; generated results should be identified as simulations. Third-party datasets and graphics assets remain subject to their original licenses and usage restrictions, including in any redistribution.

AI USE STATEMENT

We used generative AI tools, including OpenAI ChatGPT and Codex and Anthropic Claude (Claude Code), in the following ways. *Research ideation and execution:* the tools assisted with research ideation, refinement of hypotheses and mathematical descriptions, feedback on methodology and ablation design, code development and debugging, launching and monitoring experiments, collecting metrics, producing figures from saved results, and interpreting experimental results. *Retrieval and discovery:* they helped find and summarize related work and check citations against the source papers. *Drafting and polishing:* they drafted sections of the paper, including text, LaTeX equations, tables and captions, and helped revise and polish the writing. *Synthetic datasets:* they helped write and run the scripts that generate our synthetic data, namely the pretraining corpus of analytic constitutive responses and the PhysON-Validation trajectories and renderings. The data themselves are produced by our MPM simulator with analytic constitutive laws and by a path tracer, not by a generative model, and the PhysON-Validation trajectories were checked by replaying the analytic laws. Our workflow included checking suggestions against source papers and implementation code, inspecting experiment logs, saved metrics and rendered media, and compiling the manuscript to check formatting. AI-generated explanations are not themselves experimental evidence. The authors take full responsibility for the final methods, data, results, citations, and text, including content produced with AI assistance.

REFERENCES

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*, volume 37, pp. 75035–75063. Curran Associates, Inc., 2024. doi: 10.52202/079017-2388. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/89379d5fc6eb34ff98488202fb52b9d0-Abstract-Conference.html.

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*, volume 37, pp. 65643–65669. Curran Associates, Inc., 2024. doi: 10.52202/079017-2097. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/78b6beab44f92adc74ac1fdb212ac3a0-Abstract-Conference.html.

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. URL https://openaccess.thecvf.com/content/CVPR2025/html/Chen_Vid2Sim_Generalizable_Video-based_Reconstruction_of_Appearance_Geometry_and_Physics_for_CVPR_2025_paper.html.

Hsiao-yu Chen, Edith Tretschk, Tuur Stuyck, Petr Kadlecek, Ladislav Kavan, Etienne Vouga, and Christoph Lassner. Virtual Elastic Objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15827–15837, 2022. URL https://openaccess.thecvf.com/content/CVPR2022/html/Chen_Virtual_Elastic_Objects_CVPR_2022_paper.html.

Berthy T. Feng, Alexander C. Ogren, Chiara Daraio, and Katherine L. Bouman. Visual Vibration Tomography: Estimating Interior Material Properties From Monocular Video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16231–16240, 2022. URL https://openaccess.thecvf.com/content/CVPR2022/html/Feng_Visual_Vibration_Tomography_Estimating_Interior_Material_Properties_From_Monocular_Video_CVPR_2022_paper.html.

Sara Fridovich-Keil, Giacomo Meanti, Frederik Rahbæk Warburg, Benjamin Recht, and Angjoo Kanazawa. K-Planes: Explicit Radiance Fields in Space, Time, and Appearance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12479–12488, 2023. URL https://openaccess.thecvf.com/content/CVPR2023/html/Fridovich-Keil_K-Planes_Explicit_Radiance_Fields_in_Space_Time_and_Appearance_CVPR_2023_paper.html.

Genesis Authors. Genesis: A Generative and Universal Physics Engine for Robotics and Beyond, 2024. URL https://github.com/Genesis-Embodied-AI/genesis-world.

Shanyan Guan, Huayu Deng, Yunbo Wang, and Xiaokang Yang. NeuroFluid: Fluid Dynamics Grounding with Particle-Driven Neural Radiance Fields. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 7919–7929. PMLR, 2022. URL https://proceedings.mlr.press/v162/guan22a.html.

Yuanming Hu, Yu Fang, Ziheng Ge, Ziyin Qu, Yixin Zhu, Andre Pradhana, and Chenfanfu Jiang. A Moving Least Squares Material Point Method with Displacement Discontinuity and Two-Way Rigid Body Coupling. *ACM Transactions on Graphics*, 37(4):150:1–150:14, 2018.

Tianyu Huang, Haoze Zhang, Yihan Zeng, Zhilu Zhang, Hui Li, Wangmeng Zuo, and Rynson W. H. Lau. DreamPhysics: Learning Physics-Based 3D Dynamics with Video Diffusion Priors. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 3733–3741, 2025. doi: 10.1609/aaai.v39i4.32389. URL https://ojs.aaai.org/index.php/AAAI/article/view/32389.

Yi-Hua Huang, Yang-Tian Sun, Ziyi Yang, Xiaoyang Lyu, Yan-Pei Cao, and Xiaojuan Qi. SC-GS: Sparse-Controlled Gaussian Splatting for Editable Dynamic Scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4220–4230, 2024. URL https://openaccess.thecvf.com/content/CVPR2024/html/Huang_SC-GS_Sparse-Controlled_Gaussian_Splatting_for_Editable_Dynamic_Scenes_CVPR_2024_paper.html.

Chenfanfu Jiang, Craig Schroeder, Joseph Teran, Alexey Stomakhin, and Andrew Selle. The Material Point Method for Simulating Continuum Materials. In *ACM SIGGRAPH 2016 Courses*, pp. 24:1–24:52, 2016.

Hanxiao Jiang, Hao-Yu Hsu, Kaifeng Zhang, Hsin-Ni Yu, Shenlong Wang, and Yunzhu Li. PhysTwin: Physics-Informed Reconstruction and Simulation of Deformable Objects from Videos. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 7219–7230, 2025. URL https://openaccess.thecvf.com/content/ICCV2025/html/Jiang_PhysTwin_Physics-Informed_Reconstruction_and_Simulation_of_Deformable_Objects_from_Videos_ICCV_2025_paper.html.

Gergely Klár, Theodore Gast, Andre Pradhana, Chuyuan Fu, Craig Schroeder, Chenfanfu Jiang, and Joseph Teran. Drucker-Prager Elastoplasticity for Sand Animation. *ACM Transactions on Graphics*, 35(4):103:1–103:12, 2016.

Xuan Li, Yi-Ling Qiao, Peter Yichen Chen, Krishna Murthy Jatavallabhula, Ming Lin, Chenfanfu Jiang, and Chuang Gan. PAC-NeRF: Physics Augmented Continuum Neural Radiance Fields for Geometry-Agnostic System Identification. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=tVkrbkz42vc.

Zhengqi Li, Richard Tucker, Forrester Cole, Qianqian Wang, Linyi Jin, Vickie Ye, Angjoo Kanazawa, Aleksander Holynski, and Noah Snavely. MegaSaM: Accurate, Fast and Robust Structure and Motion from Casual Dynamic Videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10486–10496, 2025. URL https://openaccess.thecvf.com/content/CVPR2025/html/Li_MegaSaM_Accurate_Fast_and_Robust_Structure_and_Motion_from_Casual_CVPR_2025_paper.html.

Yuchen Lin, Chenguo Lin, Jianjin Xu, and Yadong Mu. OmniPhysGS: 3D Constitutive Gaussians for General Physics-Based Dynamics Generation. In *International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=9HZtP6I5lv.

Chunjiang Liu, Xiaoyuan Wang, Qingran Lin, Albert Xiao, Haoyu Chen, Shizheng Wen, Hao Zhang, Lu Qi, Ming-Hsuan Yang, Laszlo A. Jeni, Min Xu, and Yizhou Zhao. MOSIV: Multi-Object System Identification from Videos. In *International Conference on Learning Representations*, volume 2026, pp. 66204–66234, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/6cb00ce1a21a090a3dae04cebebd8341-Abstract-Conference.html.

Fangfu Liu, Hanyang Wang, Shunyu Yao, Shengjun Zhang, Jie Zhou, and Yueqi Duan. Physics3D: Learning Physical Properties of 3D Gaussians via Video Diffusion. *arXiv preprint arXiv:2406.04338*, 2024a. URL https://arxiv.org/abs/2406.04338.

Shaowei Liu, Zhongzheng Ren, Saurabh Gupta, and Shenlong Wang. PhysGen: Rigid-Body Physics-Grounded Image-to-Video Generation. In *European Conference on Computer Vision (ECCV)*, pp. 360–378, 2024b. doi: 10.1007/978-3-031-73007-8_21. URL https://stevenlsw.github.io/physgen/.

Jonathon Luiten, Georgios Kopanas, Bastian Leibe, and Deva Ramanan. Dynamic 3D Gaussians: Tracking by Persistent Dynamic View Synthesis. In *2024 International Conference on 3D Vision (3DV)*, 2024. doi: 10.1109/3DV62453.2024.00044. URL https://dynamic3dgaussians.github.io/.

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 *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 23279–23300. PMLR, 2023. URL https://proceedings.mlr.press/v202/ma23a.html.

Himangi Mittal, Peiye Zhuang, Hsin-Ying Lee, and Shubham Tulsiani. UniPhy: Learning a Unified Constitutive Model for Inverse Physics Simulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16208–16218, 2025. URL https://openaccess.thecvf.com/content/CVPR2025/html/Mittal_UniPhy_Learning_a_Unified_Constitutive_Model_for_Inverse_Physics_Simulation_CVPR_2025_paper.html.

NVIDIA. Cosmos World Foundation Model Platform for Physical AI. *arXiv preprint arXiv:2501.03575*, 2025. URL https://arxiv.org/abs/2501.03575v3.

Keunhong Park, Utkarsh Sinha, Peter Hedman, Jonathan T. Barron, Sofien Bouaziz, Dan B Goldman, Ricardo Martin-Brualla, and Steven M. Seitz. HyperNeRF: A Higher-Dimensional Representation for Topologically Varying Neural Radiance Fields. *ACM Transactions on Graphics*, 40(6), 2021. URL https://hypernerf.github.io/.

Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-NeRF: Neural Radiance Fields for Dynamic Scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10318–10327, 2021. URL https://openaccess.thecvf.com/content/CVPR2021/html/Pumarola_D-NeRF_Neural_Radiance_Fields_for_Dynamic_Scenes_CVPR_2021_paper.html.

Yi-Ling Qiao, Alexander Gao, and Ming Lin. NeuPhysics: Editable Neural Geometry and Physics from Monocular Videos. In *Advances in Neural Information Processing Systems*, volume 35, pp.

12841–12854. Curran Associates, Inc., 2022. doi: 10.52202/068431-0933. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/53d3f45797970d323bd8a0d379c525aa-Abstract-Conference.html.

Jiawei Ren, Kevin Xie, Ashkan Mirzaei, Hanxue Liang, Xiaohui Zeng, Karsten Kreis, Ziwei Liu, Antonio Torralba, Sanja Fidler, Seung Wook Kim, and Huan Ling. L4GM: Large 4D Gaussian Reconstruction Model. In *Advances in Neural Information Processing Systems*, volume 37, pp. 56828–56858. Curran Associates, Inc., 2024. doi: 10.52202/079017-1810. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/6808f2c57d9564a2639a4710e3bbd9b9-Abstract-Conference.html.

Xiyang Tan, Ying Jiang, Xuan Li, Zeshun Zong, Tianyi Xie, Yin Yang, and Chenfanfu Jiang. PhysMotion: Physics-Grounded Dynamics From a Single Image. In *2026 International Conference on 3D Vision (3DV)*, pp. 806–818. IEEE, 2026. doi: 10.1109/3DV69130.2026.00082. URL https://doi.org/10.1109/3DV69130.2026.00082.

Jianyuan Wang, Minghao Chen, Nikita Karaev, Andrea Vedaldi, Christian Rupprecht, and David Novotny. VGGT: Visual Geometry Grounded Transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5294–5306, 2025a. URL https://openaccess.thecvf.com/content/CVPR2025/html/Wang_VGGT_Visual_Geometry_Grounded_Transformer_CVPR_2025_paper.html.

Qianqian Wang, Vickie Ye, Hang Gao, Weijia Zeng, Jake Austin, Zhengqi Li, and Angjoo Kanazawa. Shape of Motion: 4D Reconstruction from a Single Video. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9660–9672, 2025b. URL https://openaccess.thecvf.com/content/ICCV2025/html/Wang_Shape_of_Motion_4D_Reconstruction_from_a_Single_Video_ICCV_2025_paper.html.

Qianqian Wang, Yifei Zhang, Aleksander Holynski, Alexei A. Efros, and Angjoo Kanazawa. Continuous 3D Perception Model with Persistent State. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10510–10522, 2025c. URL https://openaccess.thecvf.com/content/CVPR2025/html/Wang_Continuous_3D_Perception_Model_with_Persistent_State_CVPR_2025_paper.html.

Shuzhe Wang, Vincent Leroy, Yohann Cabon, Boris Chidlovskii, and Jerome Revaud. DUSt3R: Geometric 3D Vision Made Easy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20697–20709, 2024. URL https://openaccess.thecvf.com/content/CVPR2024/html/Wang_DUSt3R_Geometric_3D_Vision_Made_Easy_CVPR_2024_paper.html.

Guanjun Wu, Taoran Yi, Jiemin Fang, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Qi Tian, and Xinggang Wang. 4D Gaussian Splatting for Real-Time Dynamic Scene Rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20310–20320, 2024. URL https://openaccess.thecvf.com/content/CVPR2024/html/Wu_4D_Gaussian_Splatting_for_Real-Time_Dynamic_Scene_Rendering_CVPR_2024_paper.html.

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)*, pp. 4389–4398, 2024. URL https://openaccess.thecvf.com/content/CVPR2024/html/Xie_PhysGaussian_Physics-Integrated_3D_Gaussians_for_Generative_Dynamics_CVPR_2024_paper.html.

Jiawei Yang, Jiahui Huang, Yuxiao Chen, Yan Wang, Boyi Li, Yurong You, Maximilian Igl, Apoorva Sharma, Peter Karkus, Danfei Xu, Boris Ivanovic, Yue Wang, and Marco Pavone. STORM: Spatio-Temporal Reconstruction Model for Large-Scale Outdoor Scenes. In *International Conference on Learning Representations*, volume 2025, pp. 50446–50465, 2025a.

URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/7dee643a6c5abcdafee51496420c84dc-Abstract-Conference.html.

Yu Yang, Zhilu Zhang, Xiang Zhang, Yihan Zeng, Hui Li, and Wangmeng Zuo. PhysWorld: From Real Videos to World Models of Deformable Objects via Physics-Aware Demonstration Synthesis. *arXiv preprint arXiv:2510.21447*, 2025b. URL https://arxiv.org/abs/2510.21447.

Ziyi Yang, Xinyu Gao, Wen Zhou, Shaohui Jiao, Yuqing Zhang, and Xiaogang Jin. Deformable 3D Gaussians for High-Fidelity Monocular Dynamic Scene Reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20331–20341, 2024. URL https://openaccess.thecvf.com/content/CVPR2024/html/Yang_Deformable_3D_Gaussians_for_High-Fidelity_Monocular_Dynamic_Scene_Reconstruction_CVPR_2024_paper.html.

Chuhan Zhang, Guillaume Le Moing, Skanda Koppula, Ignacio Rocco, Liliane Momeni, Junyu Xie, Shuyang Sun, Rahul Sukthankar, Joëlle K. Barral, Raia Hadsell, Zoubin Ghahramani, Andrew Zisserman, Junlin Zhang, and Mehdi S. M. Sajjadi. Efficiently Reconstructing Dynamic Scenes One D4RT at a Time. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7382–7392, 2026. URL https://openaccess.thecvf.com/content/CVPR2026/html/Zhang_Efficiently_Reconstructing_Dynamic_Scenes_One_D4RT_at_a_Time_CVPR_2026_paper.html.

Junyi Zhang, Charles Herrmann, Junhwa Hur, Varun Jampani, Trevor Darrell, Forrester Cole, Deqing Sun, and Ming-Hsuan Yang. MonST3R: A Simple Approach for Estimating Geometry in the Presence of Motion. In *International Conference on Learning Representations*, volume 2025, pp. 82863–82886, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/ce1710ff5b66f89b0d707bf1ec2850bf-Abstract-Conference.html.

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*, pp. 388–406. Springer, 2024. doi: 10.1007/978-3-031-72627-9_22. URL https://physdreamer.github.io/.

Yizhou Zhao, Haoyu Chen, Chunjiang Liu, Zhenyang Li, Charles Herrmann, Junhwa Hur, Yinxiao Li, Ming-Hsuan Yang, Bhiksha Raj, and Min Xu. Toward Material-Agnostic System Identification from Videos. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 5944–5956, 2025. URL https://openaccess.thecvf.com/content/ICCV2025/html/Zhao_Toward_Material-Agnostic_System_Identification_from_Videos_ICCV_2025_paper.html.

Licheng Zhong, Hong-Xing Yu, Jiajun Wu, and Yunzhu Li. Reconstruction and Simulation of Elastic Objects with Spring-Mass 3D Gaussians. In *European Conference on Computer Vision*, pp. 407–423. Springer, 2024. doi: 10.1007/978-3-031-72627-9_23. URL https://www.ecva.net/papers/eccv_2024/papers_ECCV/html/271_ECCV_2024_paper.php.

Kaichen Zhou, Yuhan Wang, Grace Chen, Gaspard Beaudouin, Fangneng Zhan, Paul Pu Liang, and Mengyu Wang. PAGE-4D: VGGT-4D Perception via Disentangled Pose and Geometry Estimation. In *International Conference on Learning Representations*, volume 2026, pp. 36401–36414, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/3d55170799265c03b37993e02b71b2cc-Abstract-Conference.html.

A BACKGROUND

A.1 MATERIAL POINT METHOD

The Material Point Method (MPM) combines particles that carry material state with a background grid used to exchange momentum (Jiang et al., 2016). This separation accommodates large deformation without maintaining a deforming simulation mesh. We summarize the quadratic-kernel, affine-transfer update used in our implementation, following MLS-MPM (Hu et al., 2018). Particle i stores position $\mathbf{x}_i^n$, velocity $\mathbf{v}_i^n$, affine velocity matrix $\mathbf{C}_i^n$, and elastic deformation gradient $\mathbf{F}_i^{e,n}$ at substep n. Its mass m_i and reference volume V_i^0 are fixed. In PhysON, it additionally carries the time-independent material code $\mathbf{z}_i$. The initial state is the declared or reconstructed scene state; initial velocities need not be zero.

Stress and particle-to-grid transfer. Let h be the grid spacing, g index a grid node at $\mathbf{x}_g$, $\mathbf{r}_{ig}^n = \mathbf{x}_g - \mathbf{x}_i^n$, and w_{ig}^n be the quadratic B-spline interpolation weight. We evaluate the learned stress head on the current particle state,

$$\boldsymbol{\tau}_i^n = \mathcal{S}_{\boldsymbol{\theta}^e}(\mathbf{F}_i^{e,n}, \mathbf{C}_i^n; \mathbf{z}_i). \tag{8}$$

Here $\boldsymbol{\tau}$ is the Kirchhoff stress, rather than the Cauchy stress. For an elastic energy model, the corresponding stress convention is $\boldsymbol{\tau} = \mathbf{P}\mathbf{F}^\top$, with $\mathbf{P}$ the first Piola–Kirchhoff stress; our network predicts the response without explicitly learning an energy function. Grid mass and momentum, including the stress impulse, are accumulated as

$$m_g^n = \sum_i w_{ig}^n m_i, \tag{9}$$

$$\mathbf{p}_g^n = \sum_i w_{ig}^n \left[m_i \mathbf{v}_i^n + \left(m_i \mathbf{C}_i^n - \frac{4\Delta t}{h^2} V_i^0 \boldsymbol{\tau}_i^n \right) \mathbf{r}_{ig}^n \right]. \tag{10}$$

The reference volume and stress convention in Eq. 10 must be used together: no additional factor of $\mathbf{F}^\top$ is applied to the network's Kirchhoff-stress output.

Grid update and grid-to-particle transfer. On active grid nodes, momentum is divided by mass, external acceleration is applied, and the declared boundary rule modifies the velocity:

$$\mathbf{v}_g^{n+1} = \mathcal{B}_g\left(\frac{\mathbf{p}_g^n}{m_g^n} + \Delta t\, \mathbf{a}_g^n \right). \tag{11}$$

Here $\mathbf{a}_g^n$ includes gravity and any prescribed acceleration, and $\mathcal{B}_g$ denotes the scene's grid boundary operator. Empty nodes do not contribute. Interpolating back to the particles gives

$$\mathbf{v}_i^{n+1} = \sum_g w_{ig}^n \mathbf{v}_g^{n+1}, \qquad \mathbf{C}_i^{n+1} = \frac{4}{h^2} \sum_g w_{ig}^n \mathbf{v}_g^{n+1} (\mathbf{r}_{ig}^n)^\top, \tag{12}$$

$$\mathbf{x}_i^{n+1} = \mathbf{x}_i^n + \Delta t\, \mathbf{v}_i^{n+1}, \qquad \mathbf{F}_i^{e,\mathrm{trial},n+1} = (\mathbf{I} + \Delta t\, \mathbf{C}_i^{n+1}) \mathbf{F}_i^{e,n}. \tag{13}$$

These expressions describe the interior update; scene-specific particle boundary handling is applied where required.

Constitutive projection and differentiation. The projection head returns the elastic state for the next substep,

$$\mathbf{F}_i^{e,n+1} = \mathcal{P}_{\boldsymbol{\theta}^p}(\mathbf{F}_i^{e,\mathrm{trial},n+1}; \mathbf{z}_i). \tag{14}$$

Thus the order is current-state stress, MPM transfer and motion update, then projection. The projected state is carried forward rather than reset at observation frames. The grid couples particles with different codes, whereas the constitutive heads are queried locally. During Stage 2, derivatives through both heads and all simulated substeps connect the observation loss to the particle codes even though the head parameters are frozen. This coupling does not by itself impose a smoothness constraint on the code field.

A.2 Analytic Constitutive Laws

The pretraining corpus covers the five material families of the PAC-NeRF benchmark (Li et al., 2023), on which MASIV (Zhao et al., 2025) is also evaluated; of the five materials of MOSIV (Liu et al., 2026), it covers all but snow. Each family is defined by an analytic Kirchhoff stress $\boldsymbol{\tau}^*(\mathbf{F}^e, \mathbf{C})$ and a return map $\mathcal{Z}$ with $\mathbf{F}^{e,*} = \mathcal{Z}(\mathbf{F}^{e,\text{trial}})$, with the following parameters:

- **Elastic solids:** shear modulus μ and Lamé parameter λ, with one of three energies (neo-Hookean, fixed corotated, Hencky).
- **Newtonian fluids:** viscosity η and a bulk parameter λ, with one of four pressure laws.
- **Plasticine:** μ, λ and the yield stress σ_y at which permanent deformation begins.
- **Non-Newtonian fluids:** μ, λ, σ_y and a plastic viscosity η_p that delays the return to the yield surface.
- **Sand:** μ, λ and the Drucker–Prager friction coefficient α.

These laws generate the pretraining responses and the benchmark scenes only. PhysON never selects a family or estimates these parameters: Stage 2 optimizes material codes, and the frozen heads supply the response.

Let $\mathbf{F} = \mathbf{U}\,\mathrm{diag}(\boldsymbol{\sigma})\mathbf{V}^\top$ be the singular value decomposition, $J = \det \mathbf{F}$, $\mathbf{R} = \mathbf{U}\mathbf{V}^\top$, $\boldsymbol{\varepsilon} = \log \boldsymbol{\sigma}$ the principal Hencky strain, $\mathrm{tr}\,\boldsymbol{\varepsilon} = \sum_k \varepsilon_k$, and $\hat{\boldsymbol{\varepsilon}} = \boldsymbol{\varepsilon} - \frac{1}{3}(\mathrm{tr}\,\boldsymbol{\varepsilon})\mathbf{1}$ its deviatoric part. Singular values are clamped below at 0.05 inside the return maps. All stresses are Kirchhoff stresses evaluated on the returned $\mathbf{F}^e$.

Elasticity. The return map is the identity, $\mathcal{Z}(\mathbf{F}) = \mathbf{F}$, and the stress is one of

$$\text{neo-Hookean:} \quad \boldsymbol{\tau} = \mu\big(\mathbf{F}\mathbf{F}^\top - \mathbf{I}\big) + \lambda \log J\,\mathbf{I}, \tag{15}$$

$$\text{fixed corotated:} \quad \boldsymbol{\tau} = 2\mu(\mathbf{F} - \mathbf{R})\mathbf{F}^\top + \lambda J(J-1)\,\mathbf{I}, \tag{16}$$

$$\text{Hencky:} \quad \boldsymbol{\tau} = \mathbf{U}\,\mathrm{diag}\big(2\mu\boldsymbol{\varepsilon} + \lambda\,\mathrm{tr}(\boldsymbol{\varepsilon})\mathbf{1}\big)\mathbf{U}^\top. \tag{17}$$

Newtonian fluid. A volumetric pressure is combined with a viscous term on the affine velocity,

$$\boldsymbol{\tau} = p(J)\,\mathbf{I} + 2\eta\,\mathrm{dev}\big(\mathrm{sym}\,\mathbf{C}\big), \qquad p(J) \in \big\{\lambda(J - J^{-k}),\ \lambda(J-1),\ \lambda J(J-1),\ \lambda \log J\big\}, \tag{18}$$

with $\mathrm{sym}\,\mathbf{C} = \frac{1}{2}(\mathbf{C} + \mathbf{C}^\top)$ and $\mathrm{dev}\,\mathbf{A} = \mathbf{A} - \frac{1}{3}\,\mathrm{tr}(\mathbf{A})\mathbf{I}$. The four pressure laws are the power law with exponent k, the linear law, the law of the NCLaw water model (Ma et al., 2023), and the logarithmic law. The return map $\mathcal{Z}(\mathbf{F}) = J^{1/3}\mathbf{R}$ keeps the volume and rotation and discards shear strain.

Plasticine and non-Newtonian fluid. The stress is the Hencky law (Eq. 17), and the return map is a viscoplastic von Mises projection. With the trial deviatoric stress $\mathbf{s} = 2\mu\hat{\boldsymbol{\varepsilon}}$ and the yield excess $\delta = \|\mathbf{s}\| - \sqrt{2/3}\,\sigma_y$,

$$\mathcal{Z}(\mathbf{F}) = \mathbf{U}\,\mathrm{diag}\big(\exp\boldsymbol{\varepsilon}'\big)\mathbf{V}^\top, \qquad \boldsymbol{\varepsilon}' = \boldsymbol{\varepsilon} - \frac{\max(\delta, 0)}{2\mu\big(1 + \eta_p/(2\hat{\mu}\,\Delta t)\big)}\,\frac{\hat{\boldsymbol{\varepsilon}}}{\|\hat{\boldsymbol{\varepsilon}}\|}, \qquad \hat{\mu} = \tfrac{1}{3}\mu\textstyle\sum_k \sigma_k^2. \tag{19}$$

Plasticine uses $\eta_p = 0$, which returns the strain exactly to the yield surface as in MOSIV. Non-Newtonian fluids use $\eta_p > 0$, so the excess strain relaxes toward the surface over several substeps.

Sand. The stress is the Hencky law, and the return map is the Drucker–Prager projection of Klár et al. (2016) without cohesion. With

$$\delta\gamma = \|\hat{\boldsymbol{\varepsilon}}\| + \frac{3\lambda + 2\mu}{2\mu}\,\alpha\,\mathrm{tr}\,\boldsymbol{\varepsilon}, \qquad \alpha = \sqrt{\tfrac{2}{3}}\,\frac{2\sin\phi}{3 - \sin\phi} \tag{20}$$

for friction angle ϕ, the return map is

$$\mathcal{Z}(\mathbf{F}) = \begin{cases} \mathbf{R}, & \mathrm{tr}\,\boldsymbol{\varepsilon} \geq 0, \\ \mathbf{F}, & \delta\gamma \leq 0,\ \mathrm{tr}\,\boldsymbol{\varepsilon} < 0, \\ \mathbf{U}\,\mathrm{diag}\big(\exp(\boldsymbol{\varepsilon} - \delta\gamma\,\hat{\boldsymbol{\varepsilon}}/\|\hat{\boldsymbol{\varepsilon}}\|)\big)\mathbf{V}^\top, & \text{otherwise.} \end{cases} \tag{21}$$

B More Implementation Details

B.1 Constitutive Pretraining

Architecture and material coordinates. The default Q8 model concatenates an eight-dimensional material code with 13 deformation features for the projection head and with those features plus nine rotated strain-rate entries for the stress head. Both MLPs have five hidden layers of width 256, ELU activations, and a nine-dimensional matrix output. The Q8L64 control first maps the code through an affine lift, $\mathbf{z}_{64} = \mathbf{b} + \mathbf{B}\mathbf{z}_8$. This lift can be folded into the first linear layer: $W_s s + W_z(\mathbf{b} + \mathbf{B}\mathbf{z}_8) + b = W_s s + (W_z\mathbf{B})\mathbf{z}_8 + (b + W_z\mathbf{b})$. Consequently, this comparison concerns training parameterization and regularization, rather than an increase from eight to 64 independent material coordinates. The 200 bank entries index training material instances, not 200 constitutive heads.

Constitutive losses. In Eq. 4, both supervised terms are entrywise mean squared errors over the 3×3 matrices of a batch, divided by one global scale per term:

$$\mathcal{L}_{\text{stress}} = \frac{1}{9B}\sum_{b=1}^{B}\frac{\|\hat{\boldsymbol{\tau}}_b - \boldsymbol{\tau}_b^*\|_{\text{F}}^2}{s_\tau^2}, \quad \mathcal{L}_{\text{proj}} = \frac{1}{9B}\sum_{b=1}^{B}\frac{\|(\hat{\mathbf{F}}_b^e - \mathbf{F}_b^{e,\text{trial}}) - (\mathbf{F}_b^{e,*} - \mathbf{F}_b^{e,\text{trial}})\|_{\text{F}}^2}{s_p^2}, \tag{22}$$

where $s_\tau = 7146.7\,\text{Pa}$ and $s_p = 2.69 \times 10^{-3}$ are the root-mean-square Frobenius norms of the stress and of the projection correction over all 1.6 billion training rows. The projection term thus measures the correction applied by the return map rather than the full deformation gradient, which is close to the identity. The code terms are

$$\mathcal{L}_z = \frac{1}{B}\sum_{b=1}^{B}\|\mathbf{z}_{m(b)}\|_2, \qquad \mathcal{L}_\Sigma = \frac{1}{d^2}\left\|\tfrac{1}{M}\textstyle\sum_m \bar{\mathbf{z}}_m\bar{\mathbf{z}}_m^\top - 0.0025\,\mathbf{I}\right\|_{\text{F}}^2, \tag{23}$$

with $\bar{\mathbf{z}}_m$ the codes centered over the $M = 200$ instances, $\lambda_z = 10^{-4}$, and $\lambda_\Sigma = 10^{-3}$. Codes are centered before use, so the bank mean carries no information. No strain-rate auxiliary term is used; rate dependence is learned from the velocity-gradient input of the stress head.

B.2 Material-Field Initialization and Optimization

The $K = 32$ screening codes are the bank rows at evenly spaced indices over the 200 instances, which are ordered by family, so every family contributes candidates. Each candidate rollout covers the observed window once without gradients; its cost is a quarter to half of one optimization update. The per-particle codes start at the selected row. If the objective or its gradient becomes nonfinite after an update, that update is rejected and repeated at half the learning rate, at most eight times; the halved rate is kept. A finite gradient whose square overflows in single precision is treated the same way, since it would make Adam's second-moment estimate infinite and every later update zero. Without this safeguard, four of the eight PhysON-Validation video fits run at learning rate 10^{-2} (sand on- and off-bank, Newtonian on-bank, collision off-bank) reached a nonfinite value within 32 updates. Fits in which no nonfinite value occurs are identical with and without it. Model selection keeps the code field with the lowest observed objective seen during optimization.

All particles receive the same screened code initially, but each particle has an independent optimizable copy. The resulting field uses a single constitutive representation across the scene, rather than a separate network for each object. Screening does not assign material-family labels to particles, and a globally useful starting code need not already explain every material region. The subsequent coupled optimization is responsible for spatial differentiation.

B.3 Reconstruction, Appearance, and Rollout Evaluation

The stages optimize different quantities. Dynamic Gaussian reconstruction and volume filling establish geometry and particle correspondences. Initial-velocity fitting precedes material inference when the velocity is not provided. During material inference, the reconstructed trajectories, geometry, and initial velocity are fixed, while only particle codes are updated. After selecting the field by the observed objective, the material and simulated trajectory are fixed for appearance fitting on observed images. Ground-truth constitutive responses and phase labels are excluded from this inference path.

Table 8: **Pretraining material bank.** Laws with the number of instances, and parameter ranges over the 40 instances of each family. Moduli and yield stress in kPa, viscosities in Pa s. The sand friction coefficients correspond to friction angles of $5.2°$–$34.7°$.

Family	Laws (instances)	μ	λ	Other
Elastic	neo-Hookean (22), fixed corotated (9), Hencky (9)	6.87–575	10.9–1810	–
Newtonian	pressure: power (12), NCLaw (10), logarithmic (9), linear (9)	–	5.35–228	η 0–548
Plasticine	Hencky, von Mises (40)	41.6–152	62.2–232	σ_y 2.81–48.8
Non-Newtonian	Hencky, viscoplastic von Mises (40)	2.01–17.3	40.2–154	σ_y 0.239–2.88, η_p 0.301–5.59
Sand	Hencky, Drucker–Prager (40)	24.0–856	10.5–1250	α 0.0511–0.382

Table 9: **Computation overhead.** Wall-clock time and peak GPU memory per stage; Stage 2 rows for the on-bank plasticine block of PhysON-Validation (2,408 particles, 20 observed frames, 11 views) on an H100 shared with other jobs. Peak memory is the allocated PyTorch memory, recorded only for material inference (–: not recorded). Stage 1 time for the Q8 / Q8L64 heads, from the measured update rate without validation.

Stage	Setting	GPU	Time	Peak memory
Stage 1: pretraining	1M updates, 10,240 rows per batch	H100	4.3 / 6.8 h	–
Dynamic Gaussian reconstruction	40K iterations	H100	42 min	–
Volume filling and correspondence	10K iterations	H100	35 min	–
Screening	32 observed-window rollouts	H100	8 min	–
Material inference	60 updates	H100	55 min	7.2 GiB
Appearance	40K iterations	H100	9.5 min	–
Stage 2 total		H100	2.5 h	7.2 GiB

Future prediction starts from the same original scene state and continues through the observed and held-out intervals without inserting ground-truth intermediate states. Gradient checkpointing recomputes intermediate simulation states during the backward pass; it is a memory–compute tradeoff, not a short-window loss or a reset to observed states.

B.4 Pretraining Material Bank

Table 8 lists the 200 bank instances, 40 per family. Every instance is simulated in four scenes: a fall and a compression on a no-slip floor, a shear with free-slip walls, and a reference cube on a no-slip floor. This gives 800 trajectories (600 no-slip, 200 free-slip) and about 1.6 billion response rows, one per particle and substep. All instances have density $1000\,\mathrm{kg/m^3}$. The off-bank phases of PhysON-Validation (Table 10) use parameters absent from this bank. Seven of the 25 off-bank phases have at least one parameter outside the ranges of Table 8: plasticine phases 2–4, sand phase 0, non-Newtonian phases 1 and 2, and the elastic cube of the pool scene. The off-bank scenes therefore require extrapolation beyond the bank as well as interpolation between its instances.

B.5 Continuum Reconstruction Details

The constitutive features and output transforms are specified in Sec. 3.2. Here we give the full volume-filling and correspondence procedure summarized in Sec. 3.3.

We convert the Gaussian reconstruction into MPM particles with the volume preparation of MASIV (Zhao et al., 2025), which follows the Gaussian-Informed Continuum of GIC (Cai et al., 2024). Points of a regular lattice around the Gaussians are kept if, in every observed view, they project inside the object foreground and lie behind the rendered depth. Together with the Gaussian centers they define an occupancy volume, refined over several resolutions by $3 \times 3 \times 3$ smoothing, and lattice cells whose smoothed occupancy exceeds a threshold τ_{occ} become material points $\bar{\mathbf{X}} = [\bar{\mathbf{x}}(1), \ldots, \bar{\mathbf{x}}(Q)]^\top$; points outside the declared simulation domain are removed. The threshold sets the object boundary: a low value adds a layer of particles on every face, which at fixed mass lowers the simulated density. When the scene declares its material volume, as in our dataset, we choose τ_{occ} by comparing the filled volume with it: a single $\tau_{\text{occ}} = 0.5$, at which a boundary cell needs half of its neighborhood occupied, gives 0.98–$1.25\times$ the declared volume on all scenes except

the water pool, whose fill reaches the container walls and uses $\tau_{\text{occ}} = 0.15$ ($1.00\times$). Each particle then receives the declared volume divided by Q. All objects' particles are simulated jointly so that object interactions are retained.

We fill every observed frame in the same way and, for 10K iterations, jointly fit the canonical particle positions and a deformation network so that the advected particles match the per-frame fills under the bidirectional Chamfer distance. The advected particles give target positions $\mathbf{X}_t$ with persistent identity, so $\mathbf{X}_t$ provides point trajectories rather than an unrelated point cloud at every time step. We estimate the initial velocity as a single vector shared by all particles, fitted by 100 RAdam updates to the trajectory loss of Sec. 3.5 over the first four observed frames with a fixed starting code; when the scene declares its initial velocity field, as in our dataset, we use it directly. The reconstruction and $\mathbf{V}_0$ are kept fixed during material inference.

B.6 Training and Fitting Schedules

For Stage 1, the corpus contains 200 material instances, 40 from each of the five constitutive families, each simulated in four scenes (a fall, a shear, a compression, and a reference cube) with the analytic laws in MPM, for 800 trajectories. Every native substep contributes one response row; 204,800 rows held out from the same trajectories form the validation set. Each instance has a learned code of dimension $d = 8$, and each head is an MLP with five hidden layers of width 256 and ELU activations. We train the codes and both heads jointly with RAdam on batches of 10,240 rows, half of each batch drawn from rows above the 75th percentile of response magnitude within their trajectory, clipping each parameter group's gradient norm at 0.1. The learning rate follows a cosine schedule from 10^{-3} over the first 200K updates, then three further cosine phases from 10^{-5}, 10^{-6}, and 10^{-7} to one tenth of their start, ending at 1M updates; we use the final checkpoint. For Stage 2, we follow the dynamic reconstruction schedule of MASIV (Zhao et al., 2025) and GIC (Cai et al., 2024): 40K iterations for canonical Gaussians and their deformation network, volume filling, and 10K iterations of deformation fine-tuning on the filled particles. The initial velocity is fitted by 100 RAdam updates on the first four observed frames before material fitting, except on PhysON-Validation, where it is given as a scene condition. Screening evaluates 32 evenly spaced pretrained codes. We then optimize the per-particle codes with Adam at a learning rate of 10^{-2} (3×10^{-3} for the non-Newtonian and water-pool scenes of PhysON-Validation), for 300 updates on the public benchmarks and 60 on PhysON-Validation (40 for the water pool), with the constitutive heads frozen. We simulate a 64^3 grid with 200 substeps per video frame. The best code field by observed objective is selected; future frames are used only for evaluation. For image metrics, the fitted trajectory is then frozen and particle appearance is trained for 40K iterations on observed images. Additional settings are given in Appendix B.

C More Dataset Details

C.1 Scene Inventory and Data-Generation Protocol

The public visual system-identification benchmarks of PAC-NeRF (Li et al., 2023) and Spring-Gaus (Zhong et al., 2024), on which GIC (Cai et al., 2024) and MASIV (Zhao et al., 2025) are evaluated, contain predominantly homogeneous objects, and their material parameters are not related to the materials a learned constitutive prior was trained on. We therefore construct **PhysON-Validation**, a synthetic multiview benchmark of 14 trajectories that isolates spatially varying materials and object interactions and pairs every scene with an *off-bank* counterpart whose parameters were never seen in pretraining. Table 1 summarizes the scenes.

Scenes and materials. Five scenes contain a single heterogeneous object whose particles belong to three to five material phases of one family: an elastic cube with five stiffness phases, a plasticine block and a sand column with five phases that differ in stiffness and yield stress or friction, a Newtonian block with three viscosity phases, and a non-Newtonian sphere with three phases that differ in yield stress and viscosity. Two scenes contain interacting objects: an elastic slab colliding with a plasticine slab, and an elastic cube dropped at 1.5 m/s into a resting pool of water. The five material families are hyperelastic solids, plasticine (von Mises return), sand (Drucker–Prager), Newtonian fluid and viscoplastic non-Newtonian fluid, the analytic families of the pretraining corpus.

On- and off-bank variants. In the on-bank variant, every phase uses the exact parameters of one of the 200 material instances in the pretraining bank. The off-bank variant keeps the geometry, phase layout, initial state and all categorical laws, and changes the numeric parameters to values absent from the bank; for the single-object scenes each phase lies 0.03–2.35 (median 0.52) natural-log units from its nearest bank instance in the varied parameters. The pair therefore separates how well a method recovers materials seen during pretraining from how well it generalizes beyond them.

Simulation. Trajectories are simulated with MPM on a 65^3 grid spanning a unit cube (grid spacing $1/64$ m), with particles at 2 cm spacing, density $1000\,\mathrm{kg/m^3}$, gravity, free-slip container walls and a no-slip floor. Each of the 30 saved frames is separated by 200 substeps of $\Delta t = 1/4800$ s, i.e. 24 Hz over 1.21 s including the initial state. Constitutive stress is evaluated on the current state and the plastic return is applied after the particle-to-grid transfer, matching the pretraining corpus.

Observations and split. Every trajectory is rendered from 11 fixed calibrated cameras at 800×800 with a path tracer (32 samples per pixel). Solid and fluid surfaces are reconstructed from the particles, and sand is drawn as one sphere per particle; surface color varies smoothly with the material phase and is carried by the particles, so appearance exposes the phase layout without supervising it. Each view has a foreground mask, and the two multi-object scenes additionally provide a visible instance mask per object. Following MASIV (Zhao et al., 2025), the first 20 frames are observed and the final 10 are held out for future prediction.

Annotations. The archives retain ground-truth particle positions with persistent identities at every frame, per-particle phase, object identity and material parameters, and the deformation gradient, velocity gradient and Kirchhoff stress at each saved step. These are used only for evaluation and analysis. Reconstruction and material fitting use the images, masks, camera calibration, timestamps, the declared initial velocity field and scene settings such as the simulation domain and material volume.

C.2 PhysON-Validation: Material Phases

Table 10 lists the constitutive parameters of every phase in the on- and off-bank variant of each scene. Shear modulus μ, Lamé parameter λ and yield stress σ_y are in kPa; α is the Drucker–Prager friction coefficient; η is the Newtonian dynamic viscosity and η_p the non-Newtonian plastic viscosity (Pa s), for which λ sets the pressure response. On-bank parameters are those of pretraining bank instances. The last column is the Euclidean distance, in natural-log units over the parameters that vary within the family, from each off-bank phase to its nearest bank instance with the same constitutive laws; categorical laws and density ($1000\,\mathrm{kg/m^3}$) are shared by both variants.

C.3 PhysON-Validation: Simulation and Rendering

All 14 trajectories use the same solver settings: a 65^3 grid over a unit cube (spacing $1/64$ m), particles at 2 cm spacing ($8\times10^{-6}\,\mathrm{m^3}$ each), gravity, free-slip container walls, a no-slip floor, and 200 substeps of $\Delta t = 1/4800$ s between consecutive saved frames. The 30 saved frames span 1.21 s at 24 Hz, beginning with the initial state. Images are path-traced with 32 samples per pixel from 11 fixed calibrated cameras at 800×800; solid and fluid surfaces are reconstructed from the particle positions, and sand is rendered as one sphere per particle with radius 0.45 times the particle spacing. Surface color encodes material phase through a Gaussian-blended gradient fixed in the initial frame and advected with the particles. The two multi-object scenes also provide one visible (non-overlapping) mask per object and view.

C.4 PhysON-Validation: Initial States and Evaluation

Initial states. The on- and off-bank variants of a scene start from the same particle positions and velocities. Single objects start with a uniform velocity of 0.20–0.53 m/s; in the collision scene each slab carries its own velocity toward the other (0.54 m/s), and in the pool scene the elastic cube enters at 1.5 m/s while the water is at rest. The initial velocity field is provided to all methods.

Scene	Phase	On-bank	Off-bank	Distance
Elastic cube	0	μ 58.2, λ 188	μ 20.1, λ 70.9	0.84
	1	μ 58.8, λ 169	μ 25, λ 87.9	0.68
	2	μ 47.4, λ 153	μ 29.8, λ 105	0.43
	3	μ 47.5, λ 150	μ 34.6, λ 122	0.21
	4	μ 40.2, λ 142	μ 39.4, λ 139	0.03
Plasticine block	0	μ 120, λ 180, σ_y 35	μ 84, λ 144, σ_y 10.5	0.19
	1	μ 100, λ 150, σ_y 12	μ 65, λ 112, σ_y 3	0.49
	2	μ 100, λ 150, σ_y 10	μ 60, λ 105, σ_y 2	0.67
	3	μ 89.4, λ 134, σ_y 8	μ 49.2, λ 87.2, σ_y 1.2	1.07
	4	μ 44, λ 157, σ_y 3.85	μ 22, λ 94.1, σ_y 0.385	2.35
Sand column	0	μ 359, λ 284, α 0.381	μ 130, λ 200, α 0.02	1.35
	1	μ 162, λ 126, α 0.297	μ 222, λ 238, α 0.055	0.47
	2	μ 457, λ 304, α 0.268	μ 315, λ 275, α 0.09	0.52
	3	μ 234, λ 518, α 0.296	μ 408, λ 312, α 0.125	0.42
	4	μ 718, λ 961, α 0.283	μ 500, λ 350, α 0.16	0.28
Newtonian block	0	η 0, λ 63.3	η 20, λ 50.6	1.05
	1	η 0, λ 68.8	η 60, λ 55.1	0.42
	2	η 35, λ 75	η 150, λ 60	0.67
Non-Newtonian sphere	0	μ 5, σ_y 0.6, η_p 1	μ 10, σ_y 1.2, η_p 2	0.36
	1	μ 7, σ_y 1, η_p 2.5	μ 17.5, σ_y 2.5, η_p 6.25	0.64
	2	μ 12, σ_y 2, η_p 4	μ 36, σ_y 6, η_p 12	1.90
Elastic–plastic collision	0	μ 58.2, λ 188	μ 43.7, λ 141	0.08
	1	μ 100, λ 150, σ_y 10	μ 60, λ 105, σ_y 3	0.40
Elastic cube into water	0	μ 12.7, λ 12.7	μ 6.36, λ 8.9	0.51
	1	η 0, λ 68.8	η 250, λ 55.1	1.18

Table 10: Per-phase constitutive parameters of PhysON-Validation. Units as in the text; distance from each off-bank phase to the nearest bank instance in natural-log parameter units.

	Method	Bun	Burger	Dog	Pig	Potato	Mean
PSNR ↑	Spring-Gaus Zhong et al. (2024)	30.69	34.01	32.10	34.97	32.72	32.90
	NeuMA Cao et al. (2024)	31.27	23.78	25.61	25.40	-	26.51
	GIC Cai et al. (2024)	34.68	40.45	37.17	38.32	39.95	38.11
	MASIV Zhao et al. (2025)	40.26	**42.12**	**38.97**	**40.84**	**43.40**	**41.12**
	PhysON (ours)	**40.31**	38.82	38.71	40.09	40.79	39.74
SSIM ↑	Spring-Gaus Zhong et al. (2024)	0.992	0.994	0.994	0.996	0.992	0.994
	NeuMA Cao et al. (2024)	0.994	0.993	0.995	0.995	-	0.994
	GIC Cai et al. (2024)	0.995	**0.997**	**0.997**	0.997	0.996	0.996
	MASIV Zhao et al. (2025)	**0.998**	**0.997**	**0.997**	**0.998**	**0.998**	**0.997**
	PhysON (ours)	0.997	**0.997**	**0.997**	0.997	0.997	**0.997**

Table 11: **Observable state simulation on Spring-Gaus Real captures.** Image metrics on the observed frames of the five real captures, which provide no ground-truth particles.

Evaluation. Chamfer distance is computed over all ground-truth particles and all simulated particles at each saved frame, as the mean of the two directed mean squared nearest-neighbor distances, and reported in $10^3\,\mathrm{mm}^2$; segment values average the per-frame Chamfer distances over frames 1–19 (observed) or 20–29 (future). PSNR and SSIM average all 11 views. Because reconstructed particles do not correspond one-to-one to the ground-truth particles, indexed position errors are reported only for position-based analyses such as Table 6.

C.5 Public Benchmarks and Spring-Gaus Real Captures

PAC-NeRF and Spring-Gaus provide the public benchmark scenes described in Sec. 5.1. The five real Spring-Gaus captures do not provide ground-truth particle trajectories, so Table 11 reports image metrics rather than ground-truth geometric errors. A distance to reconstructed particle targets

measures agreement with reconstruction and must not be substituted for a distance to ground-truth particles.

PAC-NeRF. The benchmark of Li et al. (2023) contains 45 synthetic sequences, ten each of Newtonian fluid, non-Newtonian fluid, elastic and plasticine objects and five of sand, each rendered from 11 fixed calibrated cameras with ground-truth particles from the generating MLS-MPM simulation. Every sequence has 15 frames at 24 Hz, the initial state and 14 later frames, so a clip spans about 0.6 s under gravity. Following the MASIV configuration, objects are reconstructed inside a $1.0 \times 0.9 \times 1.0$ m region and simulated on a 64^3 grid with 6.25 cm spacing over a 4 m box, with a sticky ground plane at $y = 0$ and 200 substeps per frame. We fit on all frames and report the Chamfer distance over them (Table 2), the protocol of MASIV Table 1. The material family is part of the benchmark annotation; PAC-NeRF and GIC use it to select their constitutive model, whereas PhysON does not receive it.

Spring-Gaus. The synthetic subset of Zhong et al. (2024) has seven elastic objects (torus, cross, cream, apple, paste, chess, banana) observed by 10 cameras; each sequence is 30 frames at 33.3 Hz (frame interval 0.03 s), of which the first 20 are observed and the last 10 are held out for the future-prediction table, following MASIV. Objects are reconstructed inside declared regions of 0.2–0.7 m per side and simulated on the same 64^3 grid over a 4 m box, with a no-slip ground plane and again 200 substeps per frame. The five real captures provide no particles and are scored by PSNR and SSIM on the observed frames (Appendix C.5).

Initial velocity. Neither public benchmark declares the initial velocity, so it is estimated from the first three to four observed frames (Sec. 3.3); on PhysON-Validation it is a declared scene condition and is given to every method.

Metrics on the public benchmarks. Chamfer distance is computed as on PhysON-Validation (Appendix C), between all simulated particles and the benchmark's ground-truth particles at each frame, averaged over the frames of the window, in $10^3\,\mathrm{mm}^2$. The frame-0 value contains only the reconstruction error and is therefore a floor shared by all methods that fit the same reconstruction. PSNR and SSIM are computed on the held-out views against the rendered ground truth. All baseline numbers on the public benchmarks are taken from MASIV (Zhao et al., 2025), Tables 1–3; only PhysON was run by us.

D More Experiments

D.1 Ablation Studies

Initialization and spatial degrees of freedom. Table 6 separates three effects on position-supervised elastic scenes: the starting material code, gradient refinement, and spatial variation. The zero-update screened code tests selection alone; the shared-code control tests whether one spatially constant response is sufficient; the per-particle rows allow local responses to differ. Random and screened starts at the same learning rate isolate initialization more directly than comparisons that also change the learning rate. These analyses use observed particle positions and do not include errors from the video reconstruction stage.

Constitutive coverage and oracle head diagnostics. Table 7 fits codes using ground-truth constitutive responses on observed-frame states and evaluates them on held-out-frame states. It probes the frozen representation's response coverage, not whether those codes can be recovered from video or positions. Likewise, the analytic-head substitutions in Table 12 use ground-truth material laws to diagnose stress and projection errors. They are oracle controls, not unknown-material inference methods. In the same video-based plasticine ablation, the random start gives a lower future CD than the screened start (0.156 against 0.207, and 0.163 against 0.312 at the final frame), although the screened start fits observed frames 1–19 slightly more closely (0.122 against 0.129); the two runs also use different learning rates (10^{-3} and 10^{-2}). Screening is therefore not uniformly better in video-based fitting. Neither analysis alone establishes closed-loop rollout accuracy on states outside the evaluated trajectories.

Table 12: **Ablations on PhysON-Validation plasticine (on-bank).** Held-out frames 20–29; CD in $10^3\,\mathrm{mm}^2$, PSNR and SSIM over all views. The analytic-head rows replace one frozen learned head by the ground-truth law of each particle's material; they share the random start of the second row and are latent analyses, not unknown-material results. All rows: 60 updates with frozen Q8 heads, at learning rate 10^{-2} for the screened start and 10^{-3} for the random starts, as in the runs. The reconstructed initial state alone has CD 0.082.

Variant	CD future↓	CD final↓	PSNR future↑	SSIM future↑
PhysON (screened start)	0.207	0.312	32.77	0.988
Random start instead of screening	0.156	0.163	33.62	0.990
Analytic stress + learned projection	0.131	0.135	33.60	0.990
Analytic projection + learned stress	0.156	0.161	33.58	0.990

Table 13: **Permuting the fitted codes among particles** (on-bank elastic cube, random start, 300 updates). Permuted rows: mean ± sample standard deviation over three random permutations, replayed without refitting.

		Observed		Future		Final	
Heads	Codes	L1	CD	L1	CD	L1	CD
Frozen	fitted	15.81	0.423	59.41	4.298	81.98	8.916
	permuted	21.52±0.65	0.656±0.025	75.28±1.35	8.463±0.456	99.37±1.64	16.750±0.785
Fine-tuned	fitted	9.93	0.244	22.29	0.916	36.14	1.758
	permuted	16.35±0.07	0.392±0.011	53.33±2.83	2.804±0.350	60.82±5.16	3.569±0.448

Geometric and silhouette supervision. The two terms in Eq. 7 constrain different observations: correspondence-aware geometry supervises particle motion, while silhouettes supervise projected boundaries. Their combination follows the motion and silhouette guidance of MASIV (Zhao et al., 2025), applied here to particle codes rather than constitutive-network weights. A silhouette can constrain the outline without determining internal particle motion; conversely, trajectory fitting inherits errors in the reconstructed correspondences. Following the supplementary material of MASIV, we report three further ablations. All use the position-based setting of Table 6: the on-bank elastic cube (five stiffness layers, 1,000 particles), fitted to the ground-truth positions of frames 1–19 and evaluated on frames 20–29 by continuing the rollout from the initial state. All fits start from the bank-statistics random start of Table 6 with learning rate 10^{-3}, and use 10^{-5} for the heads when these are fine-tuned. Screening is not used, so these ablations examine fits that remain far from the ground truth. L1 is the mean indexed position error in mm and CD is in $10^3\,\mathrm{mm}^2$; final is frame 29.

Spatial arrangement of the codes. We take the code fields fitted in Table 6 with frozen and with fine-tuned heads, permute their codes randomly among the particles, and replay them without refitting; heads, initial state and all other settings are unchanged. Permutation raises the future L1 from 59.4 to 75.3 mm with frozen heads and from 22.3 to 53.3 mm with fine-tuned heads (Table 13). Even these poor fits therefore place their codes according to position in the object. The permutation test does not show that the fitted field recovers the true layers.

New initial velocities. We replay the same two fitted fields from the original initial positions after adding $(0, -0.4, 0)$ m/s (downward) or $(0.35, 0, 0)$ m/s (lateral) to the initial velocity. They are compared with ground truth re-simulated from the same altered state by the analytic laws; the material layout, particle identities, solver and boundaries are unchanged. None of the new motion is observed, and the windows 1–19 and 20–29 are kept only to separate early from late errors. With the downward change, both fields keep the error level of their original future prediction: late L1 is 52.8 mm against 59.4 mm with frozen heads and 17.7 mm against 22.3 mm with fine-tuned heads (Table 14). With the lateral change, the fine-tuned field, which predicted the original motion better, is worse than the frozen one at the final frame (59.1 against 17.7 mm). Fine-tuning the heads improves the fit to the observed motion but does not transfer consistently to new motion. This is one reason we keep the heads frozen. The test covers one material layout and two perturbations.

Table 14: **Replaying fitted code fields under new initial velocities** without refitting (on-bank elastic cube). Ground truth is re-simulated from the altered initial state. The whole motion is unobserved; early and late are frames 1–19 and 20–29.

		Early		Late		Final	
Added velocity	Heads	L1	CD	L1	CD	L1	CD
Downward $(0, -0.4, 0)$	frozen	18.54	0.603	52.84	3.242	78.94	8.533
	fine-tuned	12.86	0.351	17.73	0.661	26.14	1.263
Lateral $(0.35, 0, 0)$	frozen	20.03	0.507	24.04	0.618	17.68	0.193
	fine-tuned	17.76	0.433	26.42	1.033	59.05	4.075

Table 15: **Single versus multiple random starts at a budget of 100 updates** (on-bank elastic cube, frozen heads). Selection among starts uses the observed objective only.

		Observed		Future		Final	
Strategy	Selected run	L1	CD	L1	CD	L1	CD
One start, 100 updates	100	18.48	0.548	68.99	7.228	92.88	14.833
Four starts $\times$ 10, best + 60	70	19.27	0.593	71.36	8.166	95.70	16.695

Multiple random starts at equal budget. Screening picks the start from 32 gradient-free rollouts of bank codes. An alternative spends part of the update budget on several random starts. With 100 updates in total, we compare one random start optimized for 100 updates with four random starts (different seeds) optimized for 10 updates each, of which the one with the lowest observed objective is continued for 60 more. The multi-start run is slightly worse on every segment, with a future L1 of 71.4 against 69.0 mm (Table 15). All four starts lie near the bank mean, whereas screening evaluates codes of every family and reaches 8.92 mm future L1 before any update (Table 6). Neither run has converged after 100 updates.

D.2 Sensitivity Analysis

MOSIV measures robustness to reconstruction error by adding noise to the reconstructed Gaussians. In our pipeline, the geometry passed to the simulator is set by volume filling (Sec. 3.3), where the occupancy threshold τ_{occ} decides how far the particle volume extends beyond the observed surface. Table 16 varies τ_{occ} on the on-bank reconstructions of PhysON-Validation. For each value it reports the filled volume relative to the declared material volume, and the frame-0 Chamfer distance between the filled particles and the ground-truth particles. Only the volume ratio selects τ_{occ}; the Chamfer distance is a check and uses ground truth that is unavailable at inference. A low threshold adds a boundary layer on every face: at 0.05 the volume grows by up to $2.6\times$, and the Chamfer distance is 1.5–3.3 times its value at 0.5. At $\tau_{\text{occ}} = 0.5$ the six scenes other than the pool fill 0.98–$1.25\times$ their volume, and their Chamfer distance is at most $0.006 \times 10^3\,\text{mm}^2$ above the lowest value over the five thresholds. At 0.7, all of them except the non-Newtonian sphere are under-filled (0.71–$0.84\times$). The fill of the water pool reaches the container walls, and its volume changes much less with the threshold: the ratio is 1.05, 1.02, 1.00, 0.98 and 0.94 at 0.05, 0.10, 0.15, 0.20 and 0.30, so the pool uses 0.15.

Fitted codes compensate for geometric error. With an earlier threshold of 0.05, the elastic cube was filled with 2,158 particles spanning $0.24 \times 0.26 \times 0.24$ m, whereas the ground truth has 1,000 particles spanning 0.18 m. The code field fitted to the video with this geometry was mapped to the ground-truth particles by nearest neighbor (mean distance 11.9 mm) and replayed from the ground-truth initial state. Its observed, future and final L1 are 10.5, 73.2 and 118.3 mm. The screened code from which that fit started gives 3.47, 8.92 and 4.26 mm on the same geometry (Table 6). During the fit, the codes moved away from the screened code to reproduce the observed motion with the larger and therefore less dense object, and on the true geometry they predict that motion poorly. Codes fitted to video thus absorb errors of the reconstructed geometry, which is why τ_{occ} is chosen by the declared volume. All video-based PhysON-Validation results in this paper use $\tau_{\text{occ}} = 0.5$, and 0.15 for the pool.

Table 16: **Sensitivity of the filled geometry to the occupancy threshold** τ_{occ} (on-bank scenes). V/V_0: filled volume over declared material volume. CD: frame-0 Chamfer distance to the ground-truth particles ($10^3\,\text{mm}^2$), reported as a check and not used for selection. We use $\tau_{\text{occ}} = 0.5$, and 0.15 for the water pool ($V/V_0 = 1.00$, CD 0.253).

	0.05		0.3		0.5		0.7		0.9	
Scene	V/V_0	CD	V/V_0	CD	V/V_0	CD	V/V_0	CD	V/V_0	CD
Elastic cube	2.16	0.344	1.73	0.256	1.22	0.159	0.76	0.159	0.70	0.174
Plasticine block	1.98	0.269	1.58	0.162	1.20	0.082	0.84	0.078	0.72	0.100
Sand column	1.92	0.348	1.61	0.269	1.19	0.181	0.80	0.178	0.74	0.190
Newtonian block	1.37	0.226	1.15	0.172	0.98	0.148	0.78	0.185	0.68	0.239
Non-Newtonian sphere	1.64	0.323	1.33	0.216	1.18	0.183	1.01	0.177	0.85	0.183
Elastic–plastic collision	2.60	0.321	1.82	0.189	1.25	0.103	0.71	0.105	0.58	0.139
Elastic cube into water	1.05	0.233	0.94	0.277	0.86	0.316	0.75	0.401	0.68	0.530

D.3 Quantitative Results

Following MOSIV, Table 17 averages the PhysON-Validation results over the seven scenes of each variant. It adds the final frame (frame 29), where prediction errors accumulate, to the observed and future segments of Tables 4 and 5.

Table 17: **Average performance on PhysON-Validation** over the seven scenes of each variant. Observed: frames 1–19; future: frames 20–29; final: frame 29. CD in $10^3\,\mathrm{mm}^2$; PSNR and SSIM over all 11 views. ‡Includes provisional values of fits that are being rerun (Table 4); pending averages lack a scene.

		Observed			Future			Final		
Variant	Method	CD↓	PSNR↑	SSIM↑	CD↓	PSNR↑	SSIM↑	CD↓	PSNR↑	SSIM↑
On-bank	MOSIV (Liu et al., 2026)	77.43	14.59	0.874	525.9	12.59	0.813	834.6	12.50	0.804
	OmniPhysGS (Lin et al., 2025)	7.821	18.79	0.915	20.71	14.96	0.881	27.32	14.78	0.880
	PhysON (ours)	0.322‡	28.73‡	0.962‡	0.673‡	25.20‡	0.949‡	1.192‡	24.57‡	0.946‡
Off-bank	MOSIV (Liu et al., 2026)	[xx]	[xx]	[xx]	[xx]	[xx]	[xx]	[xx]	[xx]	[xx]
	OmniPhysGS (Lin et al., 2025)	[xx]	[xx]	[xx]	[xx]	[xx]	[xx]	[xx]	[xx]	[xx]
	PhysON (ours)	0.412‡	29.58‡	0.969‡	0.463‡	27.25‡	0.958‡	0.461‡	27.13‡	0.956‡

D.4 Qualitative Comparisons

Observed motion and future rollouts. Following MASIV and MOSIV, we compare observed and predicted frames with the ground truth on PhysON-Validation. Figure 3 plots the per-frame errors over the whole trajectory, as in Fig. 3 of MASIV (Zhao et al., 2025). Figures 4–7 show each scene in two views: an RGB strip from dataset camera 0 at full frame, and the particles in the fixed cube shared by all scenes of the dataset. The frames 0, 6, 12, 17, 23 and 29 are spaced uniformly over the 30 frames. Frames 0–17 lie in the observed window, and frames 23 and 29 are predicted by continuing the rollout from the initial state without any reset; a white bar marks the boundary. Frames marked * are the last ground-truth frame before first contact and the first frame in contact. Contact is with the floor for single objects and between the two slabs in the collision scene, and it is detected from the ground-truth positions only to choose these display frames. All results are the screened fits with 60 updates on the reconstructions filled with $\tau_{\text{occ}} = 0.5$. MOSIV and OmniPhysGS have not yet been run on PhysON-Validation, and their rows are [xx].

On the plasticine block (Fig. 5), both variants reproduce the fall, the impact at frames 5–6 and the flattened shape after it, including the stronger spreading of the off-bank block, whose moduli and yield stresses are lower. In the predicted frames, the on-bank per-frame Chamfer distance rises from 0.138 at frame 17 to 0.312 at frame 29, whereas the off-bank value stays between 0.140 and 0.148. On the non-Newtonian sphere (Fig. 6), the simulated on-bank sphere is already visibly sheared at frames 5–6, while the ground-truth sphere is still round, and its Chamfer distance peaks shortly after impact (Fig. 3). Once the material has spread over the floor, the on-bank Chamfer distance decreases to 0.148 at frame 29. Off-bank, it stays between 0.172 and 0.191 in all shown frames. On the elastic cube (Fig. 4), the prediction follows the impact and the tilt of the cube through frame 17. Afterwards the ground-truth cube keeps rotating and turns onto another face, whereas the prediction stays close to its tilted pose of frame 17; the Chamfer distance rises to 0.490 at frame 23 and 5.036 at frame 29. In the collision scene (Fig. 7), the prediction follows the approach of the two slabs, their contact at frames 5–6 and the motion that follows through the observed frames. After frame 19, the ground-truth plasticine slab separates from the elastic slab and leaves a thin plastic strand between them, whereas the simulated slabs stay in contact. The PSNR drops from 34.5 to 29.7 dB over frames 19–22 (Fig. 3), while the Chamfer distance rises more gradually, to 0.201 and 0.194 at frames 23 and 29, against 0.103 for the reconstructed initial state.

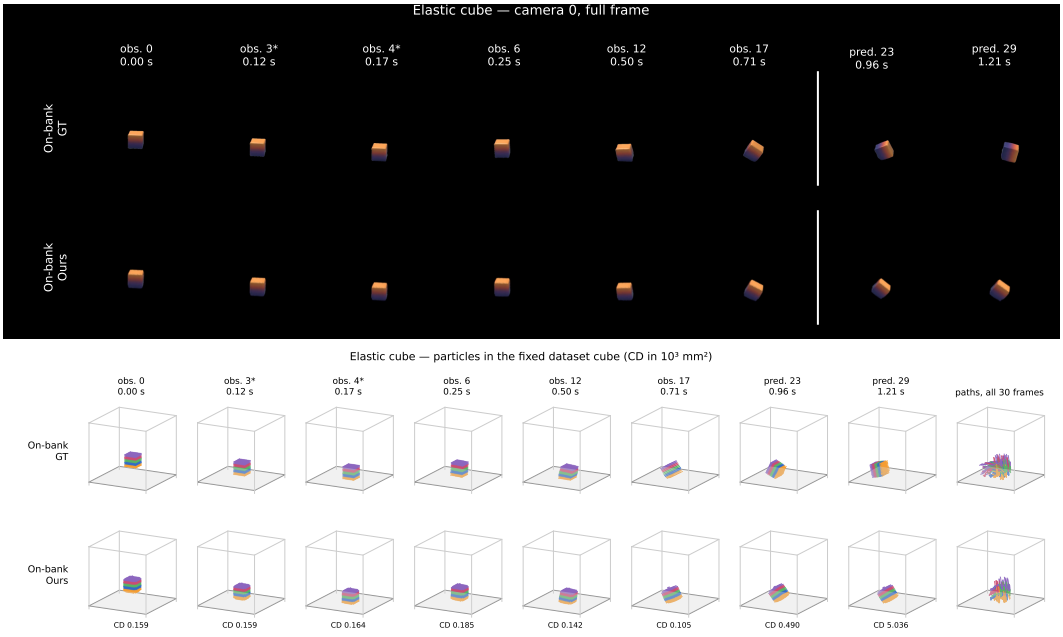


Figure 4: **Qualitative comparison on the elastic cube** (on-bank; off-bank [xx]). Top: dataset camera 0, full frame; Ours is the appearance fitted on observed images, rendered on the simulated particles. Bottom: particles in the fixed dataset cube. Ground-truth particles are colored by material phase; simulated particles take the phase of the nearest ground-truth particle at frame 0, for display only. The value under each simulated frame is its Chamfer distance ($10^3\,\mathrm{mm}^2$). The last column shows the paths of 60 randomly chosen particles over all 30 frames. Frames marked * bracket the first floor contact; the white bar separates observed from predicted frames. MOSIV and Omni-PhysGS rows: [xx].

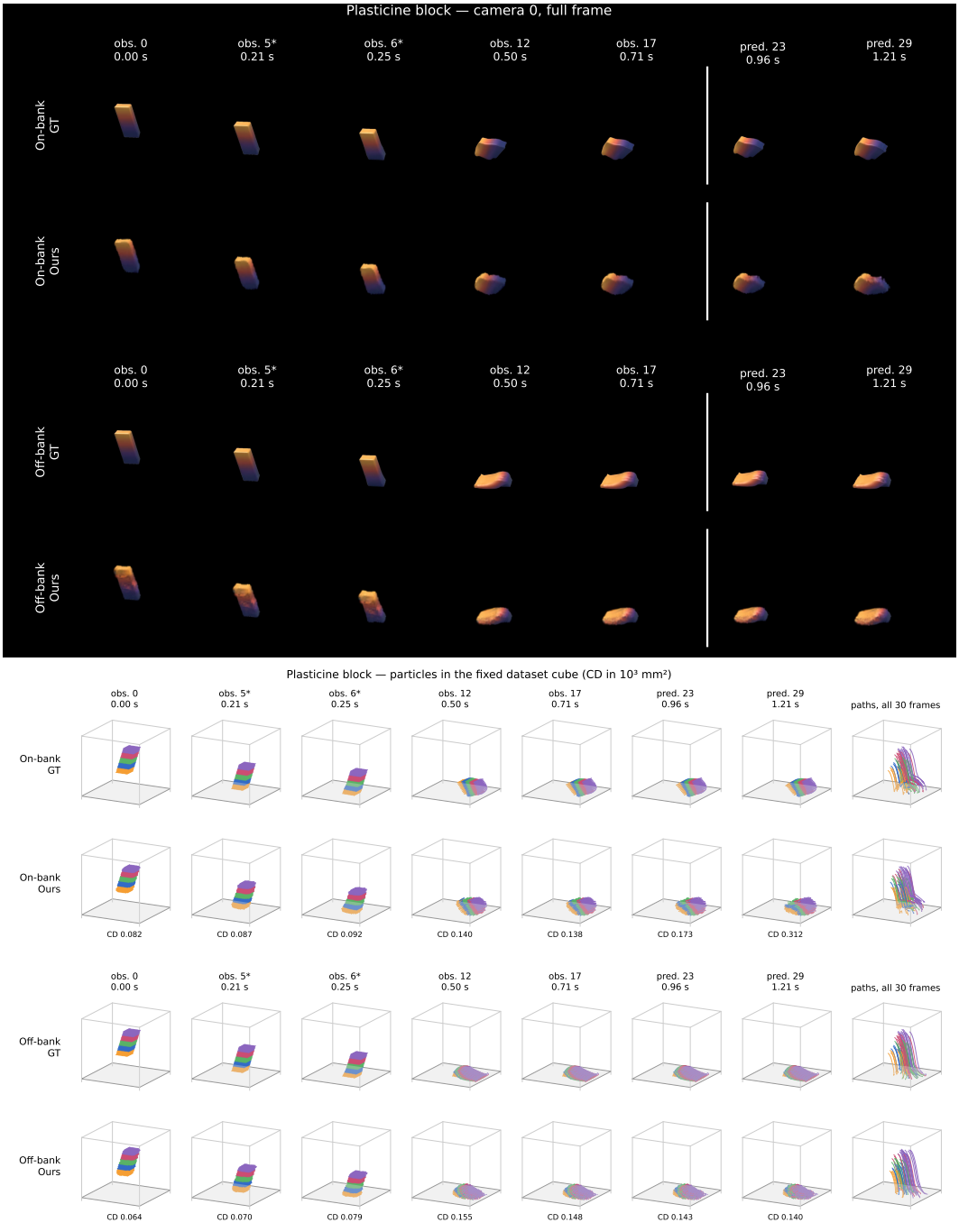


Figure 5: **Qualitative comparison on the plasticine block**, on-bank (upper rows of each panel) and off-bank (lower rows). Top: dataset camera 0, full frame. Bottom: particles in the fixed dataset cube, with Chamfer distances (10^3 mm^2) and particle paths. Layout and colors as in Fig. 4. MOSIV and OmniPhysGS rows: [xx].

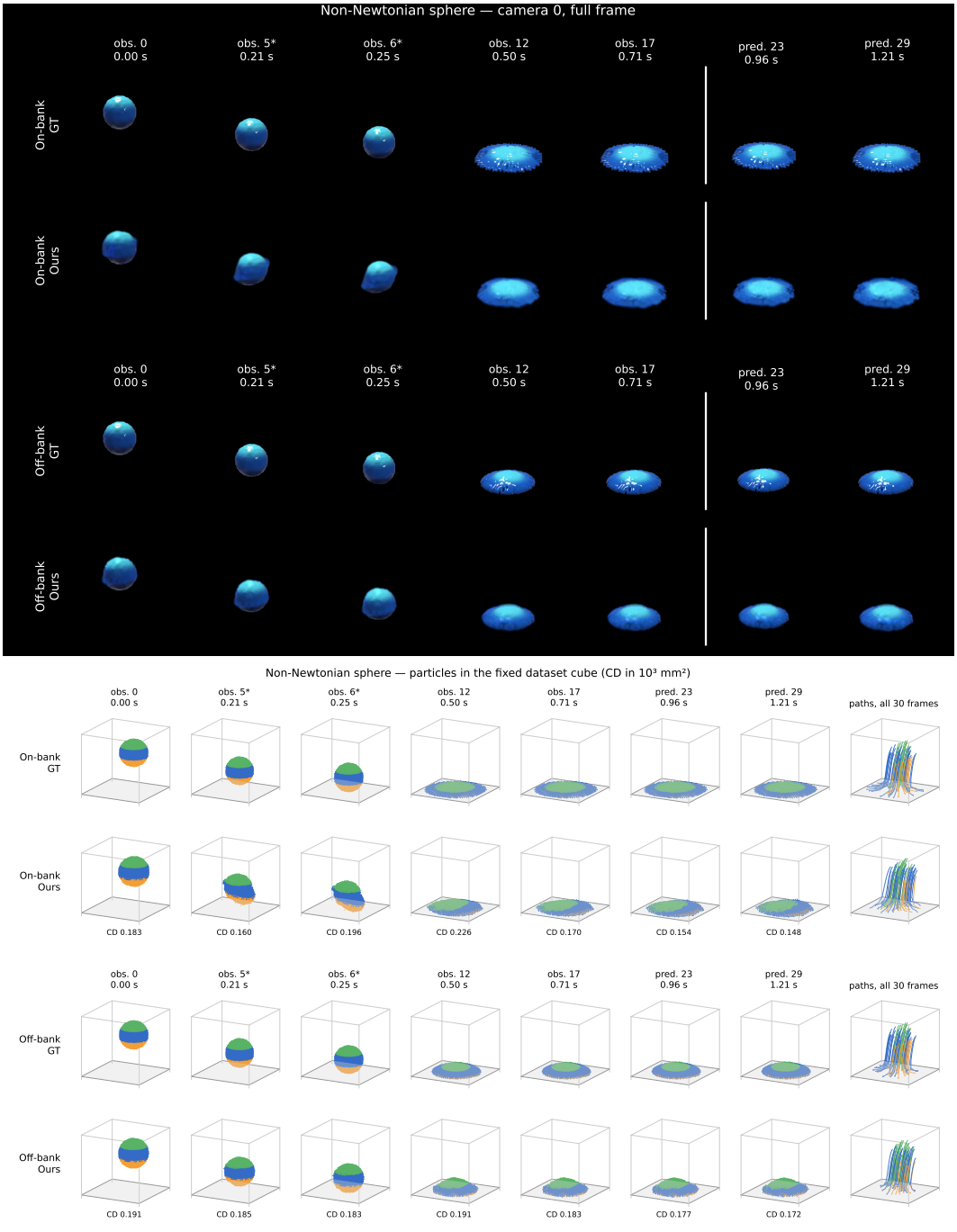


Figure 6: **Qualitative comparison on the non-Newtonian sphere**, on-bank (upper rows of each panel) and off-bank (lower rows). Top: dataset camera 0, full frame. Bottom: particles in the fixed dataset cube, with Chamfer distances (10^3 mm^2) and particle paths. Layout and colors as in Fig. 4. MOSIV and OmniPhysGS rows: [xx].

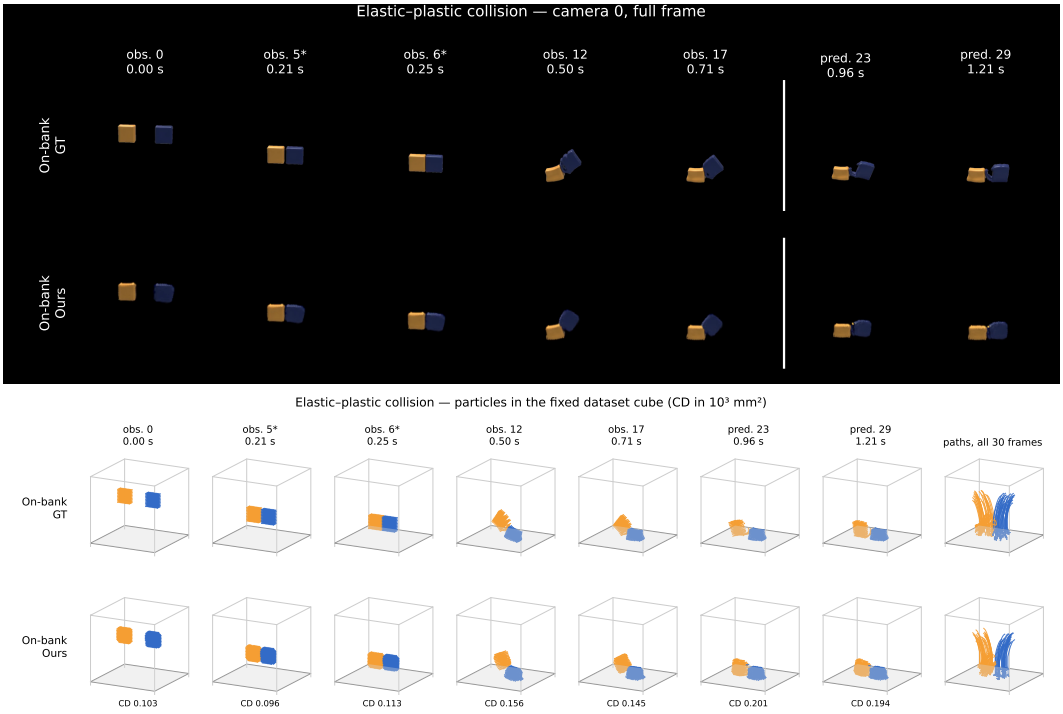


Figure 7: **Qualitative comparison on the elastic–plastic collision** (on-bank; off-bank [xx]). Top: dataset camera 0, full frame. Bottom: particles in the fixed dataset cube, with Chamfer distances ($10^3\,\mathrm{mm}^2$) and particle paths. Frames marked * bracket the first contact between the two slabs. Layout and colors as in Fig. 4. MOSIV and OmniPhysGS rows: [xx].

D.5 COUNTERFACTUAL INITIAL-STATE ROLLOUTS

Figure 8 tests whether fitted material fields remain predictive when the same scene starts differently. We replay the observed-selected code fields of three video fits, with frozen heads and without refitting, from altered initial velocities and compare them with ground truth re-simulated from the same altered states. On the plasticine block, transfer is partial. Averaged over frames 1–29, the Chamfer distance is 0.276 (downward) and 0.154 (lateral) on-bank, against 0.151 at the original velocity, and 0.133 and 0.220 off-bank, against 0.129. The largest deviation is the on-bank downward rollout, which spreads further than the ground truth and reaches 0.659 at the final frame (0.312 at the original velocity). On the elastic–plastic collision, transfer fails. Under the lateral change the ground-truth slabs bounce apart after contact, whereas ours stay in contact and tip over; the Chamfer distance peaks at 3.53 at frame 14 and averages 0.822, against 0.161 at the original velocity. Under the downward change our rollout stays below 0.25 through frame 16, deviates to 0.74 at frame 19, and becomes nonfinite at frame 23. The fitted fields thus carry over to moderate changes of a single deforming object, but not to a changed contact outcome between objects.

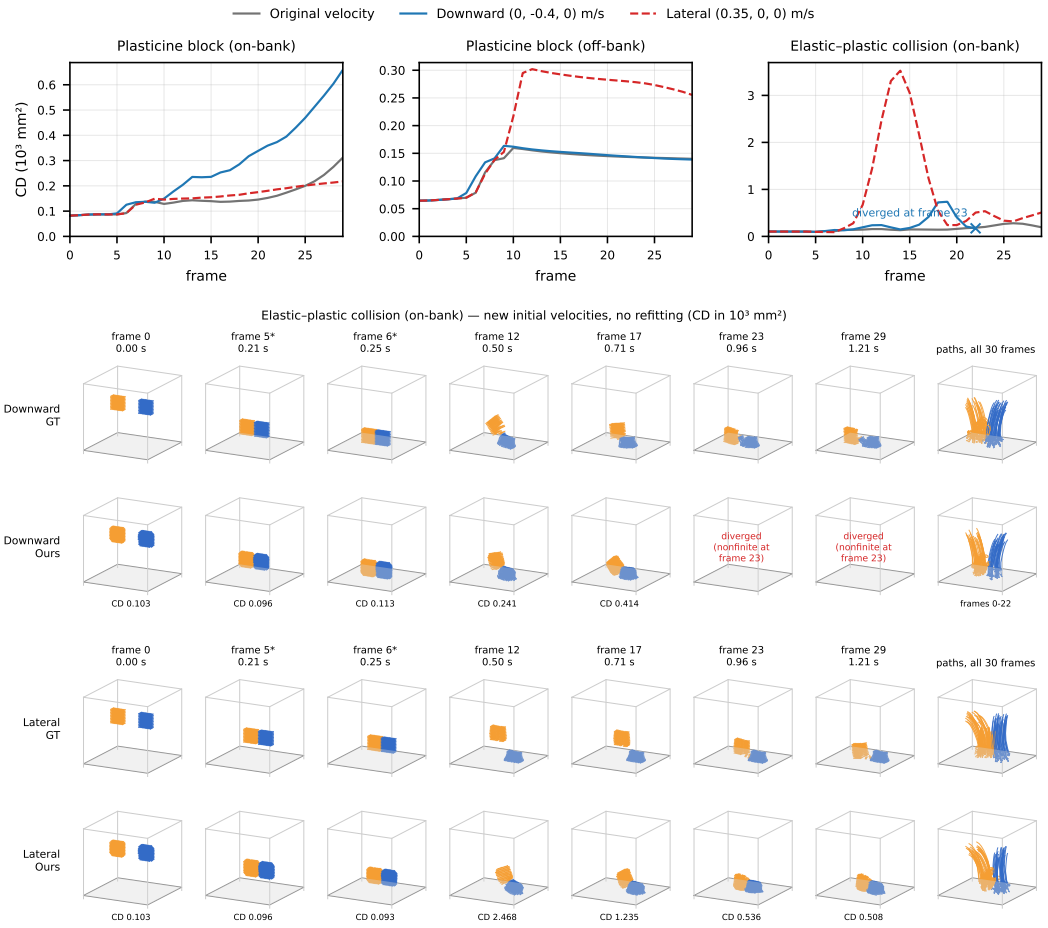


Figure 8: **New initial velocities without refitting.** The fitted material fields of three PhysON-Validation video fits are replayed from their reconstructed initial state after adding $(0, -0.4, 0)$ m/s (downward) or $(0.35, 0, 0)$ m/s (lateral) to every particle's initial velocity, and compared with ground truth re-simulated by the analytic laws from the same changed state; no frame of a changed rollout is observed by the fit. Top: per-frame Chamfer distance (10^3 mm^2) at the original and the two changed velocities. Bottom: particles of the elastic–plastic collision in the fixed dataset cube, ground truth and ours for each change, with the Chamfer distance under each simulated frame; frames marked * bracket the first contact. Our downward collision rollout becomes nonfinite at frame 23. Simulated particles take the color of the nearest ground-truth particle at frame 0, for display only.