WHAT RoR CAN WRITE UNDER PRIVILEGED SUPERVISION: DECODER EXPRESSIVITY FOR CONTINUOUS-TIME 3D GAUSSIANS

Anonymous authors
Paper under double-blind review

ABSTRACT

Open-loop continuous-time 3D Gaussian extrapolators forecast well under continuous observation, yet under intermittent reveals they still need a correctable *write* of off-screen geometric change. We run a controlled write-path *diagnostic* on NVFi Visible–Disappear–Reappear: GT-render reveals, constant-rate hidden change (translation or centroid rotation), and dense 3D attribute supervision—isolating which components of an image-space Reconcile-on-Reobservation (RoR) stack can record the change, not claiming a transferable real-camera world model. A co-evolving latent drift-rate is inert; a frozen translation-only head fails under in-place rotation (`fan`). Unfreezing the Stage-2 decoder so RoR can express quaternion updates accounts for most of the rotational sign flip vs. an honest static-memory baseline ($+7.8 \pm 1.3\%$ RoR-only at occ$=10$ over training seeds); gated typed on-manifold residuals add only a small, seed-sensitive margin at this operating point (typed $+8.3 \pm 1.7\%$), but that margin triples as the hidden rotation grows along either occlusion length or rotation rate ($+0.9 \rightarrow +2.7$ pp), so the typed channel is operating-point dependent rather than negligible. The same residual family specialized to translation remains useful for relocation ($+34.3 \pm 6.8\%$ inj-seed mean on `bat` @ occ$=10$), and adding the SO(3) channel converts `fan`'s relocation deficit from -12.5% to -1.9%, confirming that its residual error there is rotational. Photometric-only finetuning (state weight 0) collapses the flip (-13.6% at occ$=10$), and freezing the reconcile path localizes that collapse to a single operator while recovering a positive write ($+3.2\%$); a weak state weight (0.1) suffices to *learn* the write from scratch with no privileged initialization ($+11.0\%$ vs. $+12.1\%$ for full supervision), so image-only calibration of the reveal reconcile is the specific open problem, not privilege in the attribute signal.

1 INTRODUCTION

Dynamic neural scene representations—from neural radiance fields (Mildenhall et al., 2020) and deformable NeRFs (Pumarola et al., 2021; Park et al., 2021a;b) to 3D/4D Gaussian Splatting (Kerbl et al., 2023; Yang et al., 2024; Wu et al., 2024; Luiten et al., 2024)—make photorealistic novel-view synthesis of moving content routine. A parallel line learns continuous-time latent dynamics on a frozen Gaussian stage to *extrapolate* beyond the observed window (Wang et al., 2025; Hasani et al., 2022; Chen et al., 2018). These systems are strong open-loop predictors, yet under intermittent reveals they still need a *write* that records off-screen geometric change and reconciles it when the target is seen again—closer in spirit to an RSSM-style posterior update (Hafner et al., 2019; 2020) than to pure open-loop rollout.

We ask a narrower, falsifiable *diagnostic* question for ODE-GS-style continuous-time Gaussians (Figure 1): *given* image-space Reconcile-on-Reobservation (RoR) at reveals under privileged supervision, which stack component actually enables a positive write under in-place rotation—latent drift-rates, a frozen translation-only residual, decoder expressivity for quaternions, or typed on-manifold residual rates? Our backbone is a shared low-rank motion latent with a CfC prior (Hasani et al., 2022), trained on NVFi (Li et al., 2023) with a deformable-3DGS stage (Yang et al., 2024).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

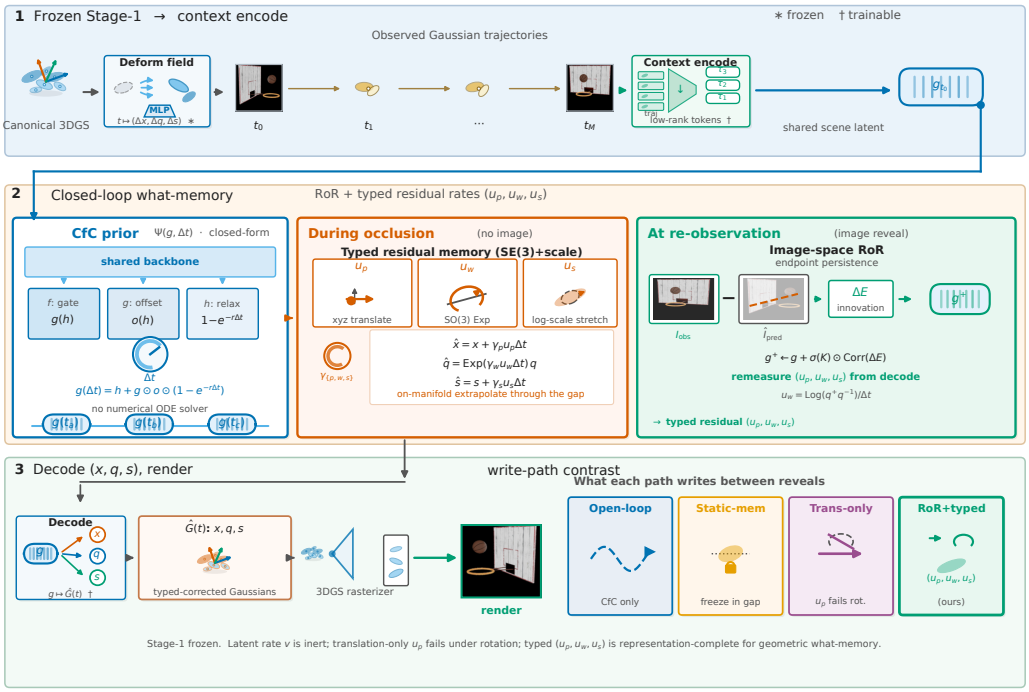

Figure 1: Write-path diagnostic stack under privileged V–D–R. **1** Stage-1 trajectories → shared latent g. **2** CfC prior rolls g; at reveals, RoR corrects g (decoder must be writable for quaternions) and optionally remeasures residual rates (u_p, u_w, u_s). **3** Decode/render; path chips contrast open-loop, static-memory, frozen translation-only, RoR with unfrozen decode, and RoR+typed.

RoR updates that latent from the innovation between a reveal image and the image rendered from the current decode. To isolate the write-path, we use a controlled protocol: reveals are *renders of the perturbed ground-truth state*; injected change is constant-rate (coherent translation, or centroid rotation of total angle $\pi/2$); and training includes dense 3D attribute MSE on hidden frames. We score per-attribute state error against an honest static-memory baseline that shares RoR anchors but freezes the target during occlusion—not image PSNR alone. Related disappear-and-reappear diagnostics in WRBench/MemoBench (Lu et al., 2026; Chen et al., 2026) motivate the structure; where-memories (Wu et al., 2025; Xiao et al., 2025), PERSIST (Garcin et al., 2026), and PersistGS (Ramlal & Zelek, 2026) address adjacent permanence questions under different observability assumptions. We do *not* claim to solve those benchmarks, nor a transferable image-only what-memory. The testable findings are: (i) unfreezing the decoder is required for RoR to write quaternions and drives most of the rotational sign flip; (ii) typed SE(3) residuals add only a small on-manifold margin at the mildest operating point, and a consistently larger one as the hidden rotation grows; (iii) a translation-specialized residual recovers relocation, and the SO(3) channel repairs the one scene where it does not; (iv) zeroing dense 3D MSE collapses the flip and the collapse localizes to the reveal reconcile operator, whereas a weak state weight (0.1) suffices to learn the write from scratch—attribute supervision is necessary, but full-weight privilege is not.

Contributions.

- A privileged write-path diagnostic on NVFi: GT-render reveals, constant-rate translation/centroid-rotation injection, dense 3D attribute supervision, and an honest static-memory baseline—explicitly *not* an image-only world-model claim.
- Mechanism attribution under that protocol: latent rates are inert; frozen translation-only fails under in-place rotation; *unfreezing the decoder* so RoR can express quaternions ac-

counts for most of the rotational sign flip; gated typed on-manifold residuals contribute a margin that scales with hidden-rotation difficulty, from $+0.9$ pp at the mildest setting to $+2.7$ pp under longer occlusion or a π rotation (Tables 5, 7).

- Closed-loop numbers with seeds: on `fan`, geodesic error vs. static reaches $+8.3 \pm 1.7\%$ (typed) / $+7.8 \pm 1.3\%$ (RoR-only) at occ=10 over training seeds (n=3); translation-specialized residuals retain relocation ($+34.3 \pm 6.8\%$ inj-seed mean on `bat`), and the SO(3) channel lifts `fan` relocation from -12.5% to -1.9%. Privilege ladder (same init): photometric-only -13.6%, weak state 0.1 $+10.6\%$, hybrid $+10.9\%$ at occ=10; trained from scratch without any privileged initialization, weak state reaches $+11.0\%$ against $+12.1\%$ for full supervision (Table 4). Image PSNR is secondary.
- A localization of the photometric-only failure: it is the reveal reconcile operator that image gradients destroy, not the decoder or the residual parameterization—freezing reconcile restores the baseline exactly and returns a positive write ($+3.2\%$), while dynamic-region pixel reweighting does not help, ruling out background dilution.

2 Related Work

Dynamic radiance fields and 3D Gaussians. Neural radiance fields (Mildenhall et al., 2020), deformable / topological NeRFs (Pumarola et al., 2021; Park et al., 2021a;b), and 3D/4D Gaussian Splatting (Kerbl et al., 2023; Yang et al., 2024; Wu et al., 2024; Luiten et al., 2024) primarily *interpolate* observed windows rather than maintain a closed-loop world state under intermittent re-observation.

Continuous-time extrapolation of dynamic Gaussians. ODE-GS-style Stage-2 models (Wang et al., 2025) freeze Stage-1 and learn latent dynamics (Neural ODEs (Chen et al., 2018), CfCs (Hasani et al., 2022), GRUs, splines) for future extrapolation. We reimplement these under a *shared* frozen Stage-1 and the same NVFi split; our contribution beyond them is a closed-loop write-path diagnostic (RoR + decoder expressivity + residual channels) that open-loop extrapolators lack. Published ODE-GS absolute scores can differ under a different Stage-1/setup (Wang et al., 2025); we do not copy those table numbers.

World models, where-memory, and occlusion permanence. RSSM/Dreamer (Hafner et al., 2019; 2020) define world models as latent dynamics with observation-conditioned posteriors. Spatial-memory / WorldMem (Wu et al., 2025; Xiao et al., 2025) store geometry or past frames for revisit consistency (where-memory), typically filtering off-screen dynamics. PERSIST (Garcin et al., 2026) maintains a persistent latent 3D state that can evolve off-screen, but does not reconcile an explicit 3DGS decode from sparse image innovations. PersistGS (Ramlal & Zelek, 2026) restores object permanence with differentiable rigid-body simulation and cameras that observe the occluded interval; we study intermittent reveals that rewrite a learned residual without a physics simulator. WRBench (Lu et al., 2026) and MemoBench (Chen et al., 2026) motivate the V–D–R structure; we study a write-path diagnostic for explicit CT-Gaussians under synthetic constant-rate change with graded attribute supervision, leaving appearance change, real-camera RoR, and fully image-only training as open.

3 Method

3.1 Shared Latent Continuous-Time Backbone

Let $\{G_i\}_{i=1}^{N}$ be Gaussians from a frozen deformable Stage-1 (Yang et al., 2024). A context window is encoded into a shared latent $g_{t_0} \in \mathbb{R}^D$. A CfC prior (Hasani et al., 2022) advances the latent for an arbitrary interval Δt:

$$g_t = \Psi_\theta(g_{t_0},\, \Delta t). \tag{1}$$

A coupled cross-attention decoder maps (g_t, q_i) to per-Gaussian attributes, where q_i is a static query embedding of Gaussian i. Because Ψ is closed-form in Δt, the same prior supports irregular reveal schedules. This open-loop path is, on its own, a strong extrapolator (Sec. 4).

3.2 IMAGE-SPACE RECONCILE-ON-REOBSERVATION

At reveal time t_r with camera C, we render $\hat{I}_{t_r}$ from the current decode and compare it to I_{t_r}. Under the hidden-change protocol, I_{t_r} is the *render of the perturbed ground-truth state* under C—a controlled oracle observation that isolates the write-path question from real-camera sensing noise. An image encoder yields $f_{\text{obs}}, f_{\text{pred}}$; the latent is corrected as

$$g_{t_r}^{+} = g_{t_r} + \sigma\big(K_\theta(g_{t_r}, \tau)\big) \odot \text{Corr}_\theta\big(f_{\text{obs}} - f_{\text{pred}},\, g_{t_r},\, \tau\big), \tag{2}$$

where τ is time since the previous correction. This is a reveal-conditioned state update toward the *updated* world (not the last visible pose). We borrow WRBench's endpoint-persistence language only as motivation: training also uses dense 3D attribute MSE on hidden frames, so the loop is not image-only.

Training freezes Stage-1 when learning the image-RoR head; for typed in-place rotation we also fine-tune the shared decoder so RoR writes can be expressed in quaternion/scale channels (Sec. 3.4).

3.3 WHY LATENT VELOCITY IS NOT ENOUGH

A natural extension maintains a latent rate v_t and rolls $g \leftarrow \Psi(g, \Delta t) + v\Delta t$, updating v from the same image innovation. Empirically this rate is inert: enabling or disabling v at evaluation yields identical predictions on our NVFi setup. The operational takeaway is that "adding velocity to the latent" does not automatically yield a useful dynamics write under our protocol.

3.4 TYPED RESIDUAL MEMORY (SE(3), OPTIONAL SCALE)

The modeled Gaussian state is 10-D: position $x \in \mathbb{R}^3$, quaternion $q \in \text{SO}(3)$, and log-scale $s \in \mathbb{R}^3$. After each reveal correction we decode the full state before/after RoR and measure *typed* residual rates

$$u_p = \frac{x^{+} - x^{-}}{\Delta t}, \quad u_w = \frac{\text{Log}(q^{+}q^{-1})}{\Delta t}, \quad u_s = \frac{s^{+} - s^{-}}{\Delta t}, \tag{3}$$

with Log the SO(3) logarithm (rotation vector). A translation-only write (u_p alone) is a strict special case; it cannot represent in-place rotation and overshoots on `fan`. During the subsequent occlusion of length Δt, each channel is gated and extrapolated in its geometry:

$$\hat{x} = x_{\text{prior}} + \gamma_p(q, \|u_p\|, \Delta t)\, u_p\, \Delta t, \tag{4}$$

$$\hat{q} = \text{Exp}\big(\gamma_w(q, \|u_w\|, \Delta t)\, u_w\, \Delta t\big)\, q_{\text{prior}}, \tag{5}$$

$$\hat{s} = s_{\text{prior}} + \gamma_s(q, \|u_s\|, \Delta t)\, u_s\, \Delta t, \tag{6}$$

with learned per-Gaussian gates $\gamma_{\{\cdot\}} \in (0, 1)$ and Exp the SO(3) exponential. Training uses mixed occlusion lengths (gaps in $\{4, \ldots, 11\}$). We call the full closed-loop stack **RoR+typed**, but ablations separate decoder unfreeze / RoR-only from the typed residual channel.

3.5 HONEST STATIC-MEMORY BASELINE

The static-memory baseline shares the same RoR corrections at reveals but freezes the dynamic-target *full state* at the post-correction anchor during occlusion. Any gain of RoR+typed over this baseline must come from inter-reveal dynamics.

4 EXPERIMENTS

4.1 SETUP

Dataset and Stage-1. Four NVFi DynamicObjects scenes (Li et al., 2023): `fallingball`, `fan`, `bat`, `whale`. Stage-1 deformable 3DGS is trained for 40k iterations on the official train split ($t \leq 0.75$). ODE / CfC / GRU / spline numbers are same-protocol reimplementations (shared Stage-1), not copied from published ODE-GS tables.

Disappear-and-reappear protocol. Closed-loop evaluation uses $\mathrm{ctx_frac} = 0.35$ of unique times as context (*distinct* from the open-loop Table 12 75% context / future split). On the horizon we reveal every $\mathrm{occ_len} \in \{4, 6, 8, 10\}$ steps when a camera exists. Dynamic targets are the top 30% Gaussians by path length. We evaluate two injection regimes: (i) *relocation*—coherent constant-velocity translation of magnitude $0.3\times$ scene extent; (ii) *in-place rotation*—coherent rotation of total angle $\pi/2$ about the target centroid (xyz via $R(p-c)+c$, quaternions via left-compose), with translation disabled. Appendix Figure 4 shows MemoBench V–D–R frames for protocol grounding only.

Metrics and evidence hierarchy. State error vs. static memory is the primary closed-loop evidence: (i) in-place geodesic SO(3) reduction (mechanism attribution under rotation); (ii) relocation RMSE reduction (translation channel). Hidden-frame image PSNR is secondary (RoR dominates). Open-loop future PSNR and HyperNeRF (Tables 12–13) are backbone sanity checks only—not the write-path claim. CfC random-access latency (Table 14) motivates closed-form Ψ under irregular reveals.

Seeds and reporting. For `fan` in-place rotation we report training-seed mean±std (n=3) across the full occlusion sweep (Table 1). For `bat` relocation under the paper gate+mix checkpoint we report injection-seed mean±std at occ=10 (n=5; $+34.3\pm6.8\%$); other reloc/rot-transfer occlusion cells remain primary-seed unless noted. Ablation Table 5 uses the primary typed checkpoint so RoR-only / xyz-only / typed share one decode.

4.2 Headline: In-Place Rotation Needs a Writable Quaternion Decode

Table 1 evaluates the *in-place rotation* protocol on `fan` (translation disabled; coherent $\pi/2$ centroid rotation)—where a frozen translation residual is structurally incomplete. A translation-only write stays near zero or negative on geodesic rotation error. With a short decoder fine-tune so RoR can express quaternion updates, the sign vs. static flips: training-seed means are occlusion-monotonic for both RoR-only ($+1.2 \to +7.8\%$) and RoR+typed ($+0.5 \to +8.3\%$). Typed on-manifold rates add only a small mean margin at occ=10 (+0.5 pp over RoR-only; +0.9 pp on the primary seed), and the margin is seed-sensitive (one seed slightly favors RoR-only). Primary-seed transfer to `bat` ($+1.1 \to +13.1\%$) and milder `whale` ($+2.8\%$ at occ=10) remains occlusion-monotonic (Table 2). The headline mechanism is decoder expressivity for quaternion RoR; typed residuals are a secondary on-manifold channel under this protocol.

The typed margin is an operating-point effect. That margin is measured where the hidden rotation is mildest. Sweeping the two axes that set how much rotation a reveal must absorb—occlusion length and injected rotation rate ρ—on the same checkpoint at evaluation only, the typed−RoR margin grows monotonically and roughly triples along *both* axes independently, from +0.9 pp to +2.7 pp at occ=16 and to +2.7 pp at $\rho=\pi$ (Appendix Table 7). RoR-only degrades once the hidden rotation exceeds what one reveal correction absorbs ($+9.1 \to +4.2\%$ at π) while the on-manifold channel decays more slowly. Retraining in-distribution at $\rho=\pi$ does *not* reproduce this (margin back to +0.6 pp), so the typed channel earns its margin in the extrapolative regime rather than from a harder training distribution.

Table 1: In-place rotation on `fan`: geodesic SO(3) error reduction vs. static (%). Translation-only is primary-seed; RoR-only / typed are training-seed mean±std (n=3) on unfrozen-decoder checkpoints. Typed−RoR mean margin at occ=10 is only +0.5 pp.

Write path	occ4	occ6	occ8	occ10
Translation-only (frozen)	−0.9	−1.5	−1.4	−1.1
RoR-only (unfrozen)	+1.2±0.8	+3.4±0.5	+4.0±0.6	+7.8±1.3
RoR+typed	+0.5±1.1	+3.0±0.2	+3.8±0.5	+8.3±1.7



Figure 2: Mid-occlusion closed-loop plate (occ=10). RGB (top) + shared-scale |pred−GT| (bottom); vermillion box = largest Ours-vs-static advantage; corner number = mean $|\Delta|$ inside the box (lower better). Open-loop drifts, static freezes, Ours darkens $|\Delta|$. Renders use geometric localization for static protection; Table 9 reports the unlocalized protocol.

Table 2: Cross-scene rotation transfer under an injected coherent rotation (primary seed unless noted). `fan`: in-place metric; training-seed means in Table 1. `bat`/`whale`: dynamic-all geodesic reduction (residual translation present).

Scene	occ4	occ6	occ8	occ10
`fan` (typed, seed mean)	+0.5	+3.0	+3.8	**+8.3**
`bat` (dyn-all, primary)	+1.1	+5.3	+8.5	**+13.1**
`whale` (dyn-all, primary)	−0.2	+0.2	+0.9	**+2.8**

4.3 Relocation: Translation Channel Remains Occlusion-Monotonic

The same typed family specialized to u_p (canonical gated translation residual) recovers coherent relocation (Table 3; Figure 3). On `fallingball`, `bat`, and `whale` the gain vs. static is positive at every length and *increases monotonically* with occlusion (primary-seed peak $+43.8\%$ on `bat`; $+34.3 \pm 6.8\%$ over injection seeds n=5 at occ=10). Under u_p alone, `fan` remains negative on relocation RMSE—not because the relocation protocol injects spin (it does not), but because `fan`'s natural motion already mixes strong in-place rotation into the dynamic-target set, so a translation residual fights a rotation-dominated scene (consistent with Table 1). The test passes: a typed head

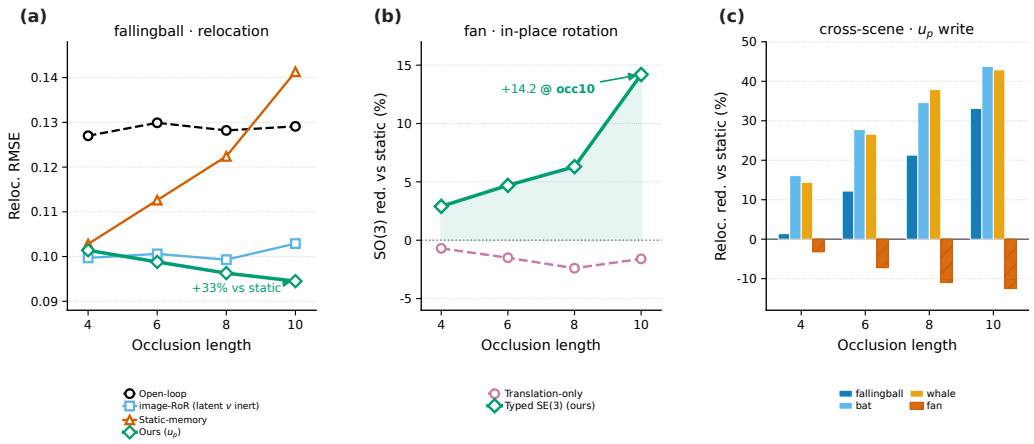

Figure 3: **(a)** Relocation RMSE vs. occlusion on `fallingball` (u_p). **(b)** In-place rotation on `fan`: geodesic SO(3) reduction vs. static—unfrozen RoR/typed vs. frozen translation-only (see Table 1 for seed means). **(c)** Cross-scene relocation RMSE reduction vs. static under u_p (grouped bars; `fan` hatched negative).

owning an SO(3)-tangent channel, evaluated under the *same* translation-injection protocol, cuts the deficit from -12.5% to -1.9% at occ=10 and reverses its slope—the xyz-only write worsens with occlusion ($-7.2 \rightarrow -12.5\%$) whereas the typed write improves ($-3.0 \rightarrow -1.9\%$). `fan`'s relocation residual is thus rotational and out of reach of a translation-only write, not evidence that reconcile fails on this scene. Appendix Table 8 widens the translational readout at occ=10.

Table 3: Relocation state RMSE: translation residual (u_p) vs. static reduction (%; primary seed). `bat` @ occ=10 over injection seeds: $+34.3 \pm 6.8\%$ (n=5).

Scene	occ4	occ6	occ8	occ10
fallingball	$+1.4$	$+12.2$	$+21.3$	$+33.1$
bat	$+16.1$	$+27.8$	$+34.6$	$+43.8$
whale	$+14.4$	$+26.6$	$+37.9$	$+42.9$
fan (u_p only)	-3.2	-7.2	-10.9	-12.5
fan (typed, u_p+u_w)	—	-3.0	-2.1	-1.9

4.4 Image PSNR Is Secondary to State Error

Hidden-frame image PSNR is secondary (Appendix Table 9). PSNR alone cannot certify a geometric write—Tables 1–3 do.

4.5 Ablations: Mechanism Ladder, then Translational Write Path

Table 5 is the mechanism table on `fan` @ occ=10 (primary typed checkpoint; rotation protocol). A frozen xyz-only head fails (-1.1%); frozen-decoder typed also fails (-2.3%). Unfreezing the decoder so RoR can express quaternion updates accounts for *most* of the sign flip: on that same checkpoint, RoR-only and unfrozen xyz-only (typed rates disabled at eval) both reach $+9.1\%$ vs. static; gated typed extrapolation adds only $+0.9$ pp to $+10.0\%$. Across training seeds the typed$-$RoR margin shrinks further (mean $+0.5$ pp; Table 1). We therefore treat **decoder expressivity as the primary write enabler** and typed residuals as a small, optional on-manifold channel—not as the named solution that "fixes" translation-only failure by itself. Ungated typed underperforms gated; random-walk injection remains positive but weaker than coherent; adding log-scale retains rotation with near-zero scale gain under this geometric protocol.

Table 4 probes supervision privilege by finetuning the same typed checkpoint under reduced state weight and/or photometric MSE on extrapolated renders. Zeroing dense 3D MSE collapses the flip (-13.6% at occ=10); keeping only 10% state weight retains a positive, occlusion-monotonic curve ($+10.6\%$), and adding photometric loss does not hurt ($+10.9\%$).

Where photometric-only actually breaks. Since our static-memory baseline is re-anchored on the reconciled decode at reveals (Section 3.4), a damaged reconcile inflates the baseline and the write together: under photometric-only finetuning the static in-place geodesic error rises from 0.55 to 0.99, which is the collapse. Freezing the reconcile path and training only the typed gates under the same image-only loss restores the baseline exactly (0.55) and returns a positive write ($+3.2\%$ at occ=10), isolating the reconcile operator—not the decoder, not the residual parameterization—as what image gradients destroy. The rate is still mis-calibrated ($+3.2\%$ against $+10.0\%$ with attribute supervision on the identical reconcile), and upweighting the pixels the hidden change moves ($20\times$) makes both variants worse (-2.7%, $+2.5\%$), ruling out background dilution. Image-only calibration of the reveal reconcile thus remains open, now narrowed to one operator.

The write does not need the privileged initialization chain. Every row above finetunes `image_ror_typed_rot`, itself the end of a 110-epoch fully-privileged chain, so a weak-supervision result could be inherited rather than learned. Training directly from the frozen backbone for 80 epochs with no privileged initialization still reproduces the flip at w_{state}=0.1 ($+2.2 \rightarrow +11.0\%$, occlusion-monotonic), against $+12.1\%$ for a paired from-scratch control at w_{state}=1. Weak attribute supervision is therefore sufficient to *learn* the rotational write, not merely to preserve one obtained under full privilege.

Table 4: Privilege ladder on `fan` (in-place rotation): geodesic SO(3) reduction vs. static (%). All rows finetune from the same primary typed checkpoint; w_{state} multiplies dense 3D attribute MSE; w_{img} is photometric MSE on extrapolated renders.

Supervision	occ4	occ6	occ8	occ10
Full (w_{state}=1, primary)	+1.9	+2.8	+3.6	+10.0
Weak state (w_{state}=0.1)	+4.7	+4.5	+5.4	+10.6
Hybrid (0.1 / w_{img}=1)	+2.0	+2.4	+4.6	+10.9
Photometric-only (w_{state}=0)	−5.8	−8.6	−10.9	−13.6
+ decoder frozen	—	−3.4	−5.2	−5.8
+ reconcile frozen (typed gates only)	—	+0.2	−1.9	+3.2
From scratch (no privileged init chain, 80 epochs)				
Weak state (w_{state}=0.1)	+2.2	+5.0	+7.6	**+11.0**
Full (w_{state}=1)	+0.9	+4.1	+9.1	**+12.1**

Table 5: Mechanism ladder on `fan` @ occ=10 (in-place rotation): geodesic SO(3) reduction vs. static (%). Rows marked † reuse the gated typed checkpoint (unfrozen decoder); RoR-only / xyz-only disable typed rates at eval. The $+0.9$ pp typed margin is the incremental claim—not the bulk of the flip.

Write / training	dyn_rot %
xyz-only residual (frozen head)	−1.1
Frozen-decoder typed	−2.3
RoR-only (unfrozen ckpt)†	+9.1
Unfrozen xyz-only (same ckpt)†	+9.1
Ungated typed	+2.4
Gated typed†	+10.0
random-walk inject	+6.4
+ log-scale channel	+10.6

For the translation channel, Table 6 compares ungated u_p vs. gate+mix; Appendix Table 10 ladders mechanisms on `fallingball`. Soft dyn-gating does not clear static-bg pollution; tables keep the unlocalized protocol.

Table 6: Cross-scene translation-channel ablation: relocation RMSE reduction vs. static (%; primary seed). Ungated = hc_coh; Gate+mix = canonical u_p (same bat checkpoint as the $+34.3 \pm 6.8\%$ inj-seed mean).

Scene	Write	occ4	occ6	occ8	occ10
fallingball	Ungated u	-11.2	-4.9	$+9.7$	$+26.4$
	Gate+mix	$\mathbf{+1.4}$	$\mathbf{+12.2}$	$\mathbf{+21.3}$	$\mathbf{+33.1}$
bat	Ungated u	$+9.4$	$+19.1$	$+27.0$	$+36.2$
	Gate+mix	$\mathbf{+16.1}$	$\mathbf{+27.8}$	$\mathbf{+34.6}$	$\mathbf{+43.8}$
whale	Ungated u	$+8.3$	$+23.6$	$+32.6$	$+39.2$
	Gate+mix	$\mathbf{+14.4}$	$\mathbf{+26.6}$	$\mathbf{+37.9}$	$\mathbf{+42.9}$
fan	Ungated u	-9.3	-10.6	-14.9	-17.2
	Gate+mix	$\mathbf{-3.2}$	$\mathbf{-7.2}$	$\mathbf{-10.9}$	$\mathbf{-12.5}$

4.6 Robustness, Backbone Sanity, and Efficiency

Drift/reveal sweeps and open-loop sanity checks are deferred (Appendix Tables 11–14; Figure 4).

5 Limitations and Discussion

Typed residuals assume approximately constant rates between reveals; large spins, nonlinear deformations, and non-constant-rate motion remain outside the tested regime—our robustness sweeps vary drift magnitude and reveal density, not the motion family. The optional log-scale channel preserves the rotation gain when added (Table 5, $+10.6$ vs. $+10.0$), but the injected log-scale change—though present in the scored ground truth (static log-scale RMSE ≈ 0.157)—is only weakly observable in the render, since a per-Gaussian anisotropic size shift barely moves overlapping-Gaussian appearance. Image-space RoR therefore cannot measure a scale rate to extrapolate, and a dedicated scale fine-tune reduces scale error by at most $+0.4\%$. Read as a *scope finding*: image-driven RoR writes SE(3) but not per-Gaussian scale. Our protocol uses synthetic constant-rate geometric change, GT-render reveals, and (in the main recipe) dense 3D supervision. Table 4 shows the dependence is graded: photometric-only finetuning fails (-13.6% at occ=10), but weak state weight 0.1 retains the flip ($+10.6\%$). The failure localizes to the reveal reconcile operator, which image gradients decalibrate; protecting it recovers a positive but under-calibrated write ($+3.2\%$), and dynamic-region pixel reweighting does not close the gap. The weak-supervision result is not inherited from the privileged initialization: trained from scratch, w_{state}=0.1 still flips the sign ($+11.0\%$ at occ=10) against $+12.1\%$ for a paired full-supervision control. We therefore claim a write-path diagnostic that needs *some* attribute supervision—not an image-only what-memory, and not a requirement for full-weight dense 3D MSE. Appearance change, real-camera closed-loop RoR, image-only calibration of the reveal reconcile, and organic MemoBench scoring remain open. Shared-latent corrections can move near-static Gaussians; we do not claim perfect self-localization.

6 Conclusion

Under a privileged NVFi V–D–R diagnostic (GT-render reveals, constant-rate injection, dense 3D supervision), recording hidden geometric change requires a reveal reconcile *and* a decode that can express the written attributes. Image-space RoR supplies the reveal update; *unfreezing the Stage-2 decoder* so RoR can write quaternions accounts for most of the rotational sign flip that a latent rate and a frozen translation-only residual failed to provide. Typed on-manifold residuals add a margin that is small at the mildest operating point ($+0.5$ pp mean at occ=10) and grows to $+2.7$ pp once the hidden rotation exceeds what one reveal absorbs, and they remain useful when specialized to translation ($+34.3 \pm 6.8\%$ inj-seed reloc on bat). Photometric-only finetuning fails (-13.6%), while weak state weight 0.1 learns the flip from scratch without privileged initialization ($+11.0\%$ vs. $+12.1\%$ full; Table 4)—so we do not claim an image-only what-memory, but the write is not an artifact of privileged supervision, and the residual barrier is localized to image-only calibration of the reveal reconcile. We claim neither open-loop SOTA nor a solution to WRBench/MemoBench, but a mechanism-level write-path diagnostic under (possibly weak) attribute supervision.

AI USE STATEMENT

(This section is **required** and does not count toward the page limit.)

In this work, we used generative AI tools for drafting and editing manuscript text, suggesting paper structure, formatting LaTeX, and assisting with implementation of training/evaluation scripts (including weak-privilege and photometric supervision probes). We have not used generative AI tools for generating synthetic datasets, formulating mathematical claims or proofs, or interpreting experimental results as unverified outputs, and the remaining required-disclosure tasks in the ICLR 2027 AI Policy are not applicable to this work. Additionally, we used generative AI tools for readability edits, literature search assistance, and code editing. We have reviewed all AI-assisted work: every scientific claim, experimental protocol, training run, and reported number was checked against local logs and regenerated metrics by the authors; AI-edited code was executed and validated on our NVFi closed-loop suite. We take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

ETHICS STATEMENT

(This section is **recommended** and does not count toward the page limit.)

This work studies geometric write-paths for continuous-time 3D Gaussian models on publicly available scene datasets (NVFi, HyperNeRF) under a synthetic constant-rate Visible–Disappear–Reappear protocol with ground-truth-render reveals. It does not involve human subjects, private personal data, or dual-use deployment claims.

REPRODUCIBILITY STATEMENT

(This section is **recommended** and does not count toward the page limit.)

Anonymous source code for the write-path diagnostic, seed sweeps, privilege-ladder probes, from-scratch supervision controls, and the negative-result root-cause probes (`run_iclr_rescue.sh`, `run_nonpriv_probe.sh`, `run_negfix_probe.sh`, `run_typedmargin_sweep.sh`, `run_fanreloc_fix.sh`, `run_scratch_weak.sh`) is in the code supplement (`gsworld/`). Protocol details are in Sec. 4; deferred tables/figures in Appendix A.

REFERENCES

Haoyu Chen, Kaichen Zhou, Hang Hua, Kaile Zhang, Jingwen Qian, Wufei Ma, Haonan Chen, Chunjiang Liu, Yizhou Zhao, Xiaoyuan Wang, et al. MemoBench: Benchmarking world modeling in dynamically changing environments. *arXiv preprint arXiv:2606.27537*, 2026.

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2018.

Samuel Garcin, Thomas Walker, Steven McDonagh, Tim Pearce, Hakan Bilen, Tianyu He, Kaixin Wang, and Jiang Bian. Beyond pixel histories: World models with persistent 3D state. *arXiv preprint arXiv:2603.03482*, 2026.

Danijar Hafner, Timothy Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. Learning latent dynamics for planning from pixels. In *International Conference on Machine Learning (ICML)*, 2019.

Danijar Hafner, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. Dream to control: Learning behaviors by latent imagination. In *International Conference on Learning Representations (ICLR)*, 2020.

Ramin Hasani, Mathias Lechner, Alexander Amini, Lucas Liebenwein, Aaron Ray, Max Tschaikowski, Gerald Teschl, and Daniela Rus. Closed-form continuous-time neural networks. *Nature Machine Intelligence*, 4(11):992–1003, 2022. doi: 10.1038/s42256-022-00556-7.

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

Jinxi Li, Ziyang Song, and Bo Yang. NVFi: Neural velocity fields for 3D physics learning from dynamic videos. In *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 34723–34751, 2023. doi: 10.52202/075280-1508.

Jinpeng Lu, Dexu Zhu, Haoyuan Shi, Linghan Cai, Guo Tang, Yinda Chen, Jie Cao, Duyu Tang, Yi Zhang, Yong Dai, and Xiaozhu Ju. Current world models lack a persistent state core. *arXiv preprint arXiv:2606.20545*, 2026.

Jonathon Luiten, Georgios Kopanas, Bastian Leibe, and Deva Ramanan. Dynamic 3d gaussians: Tracking by persistent dynamic view synthesis. In *International Conference on 3D Vision (3DV)*, pp. 800–809, 2024. doi: 10.1109/3DV62453.2024.00044.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. In *European Conference on Computer Vision (ECCV)*, pp. 405–421, 2020. doi: 10.1007/978-3-030-58452-8_24.

Keunhong Park, Utkarsh Sinha, Jonathan T. Barron, Sofien Bouaziz, Dan B. Goldman, Steven M. Seitz, and Ricardo Martin-Brualla. Nerfies: Deformable neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 5845–5854, 2021a. doi: 10.1109/ICCV48922.2021.00581.

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): 1–12, 2021b. doi: 10.1145/3478513.3480487.

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. 10313–10322, 2021. doi: 10.1109/CVPR46437.2021.01018.

Adrian Ramlal and John S. Zelek. PersistGS: Differentiable physics for object permanence in 4D gaussian splatting. *arXiv preprint arXiv:2606.03479*, 2026.

Daniel Wang, Patrick Rim, Tian Tian, Dong Lao, Alex Wong, and Ganesh Sundaramoorthi. ODE-GS: Latent ODEs for dynamic scene extrapolation with 3D gaussian splatting. *arXiv preprint arXiv:2506.05480*, 2025. We cite the method; Table 1 reports our same-protocol reimplementation under a shared frozen Stage-1, not published absolute scores.

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. doi: 10.1109/CVPR52733.2024.01920.

Tong Wu, Shuai Yang, Ryan Po, Yinghao Xu, Ziwei Liu, Dahua Lin, and Gordon Wetzstein. Video world models with long-term spatial memory. *arXiv preprint arXiv:2506.05284*, 2025.

Zeqi Xiao, Yushi Lan, Yifan Zhou, Wenqi Ouyang, Shuai Yang, Yanhong Zeng, and Xingang Pan. WorldMem: Long-term consistent world simulation with memory. *arXiv preprint arXiv:2504.12369*, 2025.

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. doi: 10.1109/CVPR52733.2024.01922.

A Deferred Tables and Figures

Table 7: Typed margin vs. hidden-rotation difficulty on `fan` (in-place geodesic SO(3) reduction vs. static, %; primary typed checkpoint, evaluation only unless noted). Upper block varies occlusion length at $\rho{=}\pi/2$; lower block varies the injected rotation rate ρ at occ=10.

Difficulty setting	RoR-only	RoR+typed	margin
occ=10, $\rho{=}\pi/2$ (Table 1 protocol)	+9.1	+10.0	+0.9
occ=12	+11.6	+12.4	+0.9
occ=14	+12.1	+13.9	+1.8
occ=16	+8.5	+11.2	**+2.7**
$\rho{=}3\pi/4$	+9.0	+11.8	**+2.8**
$\rho{=}\pi$	+4.2	+6.9	**+2.7**
$\rho{=}\pi$, retrained in-distribution	+4.2	+4.8	+0.6

Table 8: Off-screen translational breakdown at occ=10 (u_p protocol). Access: reveal-frame PSNR (dB), OL→RoR. "–": $n_{\text{inplace}} = 0$ under the motion-straightness split (all dynamic Gaussians counted as relocation on `bat`/`whale`)—*not* the geometric rotation protocol of Table 1.

Scene	reloc	in-place	deep-hid.	access PSNR	
				OL	RoR
fallingball	+33.1	+6.9	+14.3	16.76	**20.10**
bat	+43.8	–	+46.6	29.08	**31.15**
whale	+42.9	–	+46.5	28.72	**30.19**
fan	−12.5	−3.2	+1.8	21.46	**22.74**

Table 9: Hidden-frame image PSNR (dB), $\text{ctx_frac} = 0.35$. Δ_{st} / Δ_{ol}: Ours minus static / open-loop. Large Δ_{ol} with small Δ_{st} indicates RoR—not the residual write—drives most image gain.

Scene	occ	OL	Static	Ours	Δ_{ol}	Δ_{st}
fallingball	4	17.77	20.21	**20.43**	+2.66	+0.22
	6	17.68	19.82	**20.06**	+2.38	+0.24
	8	17.69	19.60	**19.90**	+2.21	+0.30
	10	17.65	19.26	**19.68**	+2.04	+0.43
fan	4	22.59	23.89	**24.14**	+1.55	+0.25
	6	22.56	23.32	**23.77**	+1.21	+0.45
	8	22.56	23.05	**23.54**	+0.98	+0.49
	10	22.54	22.74	**23.27**	+0.73	+0.53
bat	4	29.96	30.40	**32.21**	+2.26	+1.81
	6	29.89	29.70	**32.06**	+2.17	+2.36
	8	29.91	29.32	**31.76**	+1.84	+2.43
	10	29.89	28.90	**31.60**	+1.71	+2.69
whale	4	29.72	29.48	**30.41**	+0.69	+0.93
	6	29.67	29.30	**30.50**	+0.83	+1.20
	8	29.67	29.17	**30.34**	+0.67	+1.17
	10	29.63	29.00	**30.08**	+0.45	+1.08

Table 10: Translation-channel mechanism ladder on `fallingball`: relocation RMSE reduction vs. static (%). Bottom: localization tradeoff at occ=10 (same gate+mix weights).

Write path	occ4	occ6	occ8	occ10
Latent v on / off	identical (inert; $\lvert\Delta\rvert < 0.1$ pp)			
Ungated u	−11.2	−4.9	+9.7	+26.4
Gated u (no mix)	−12.1	+7.2	+25.8	+32.4
Gate+mix (u_p)	**+1.4**	**+12.2**	**+21.3**	**+33.1**
+ soft dyn-gate	+4.9	+13.2	+17.1	+28.3
Loc. @ occ10: unloc. 19.68/sbg 0.024; geom. 17.67/sbg 0.0008.				

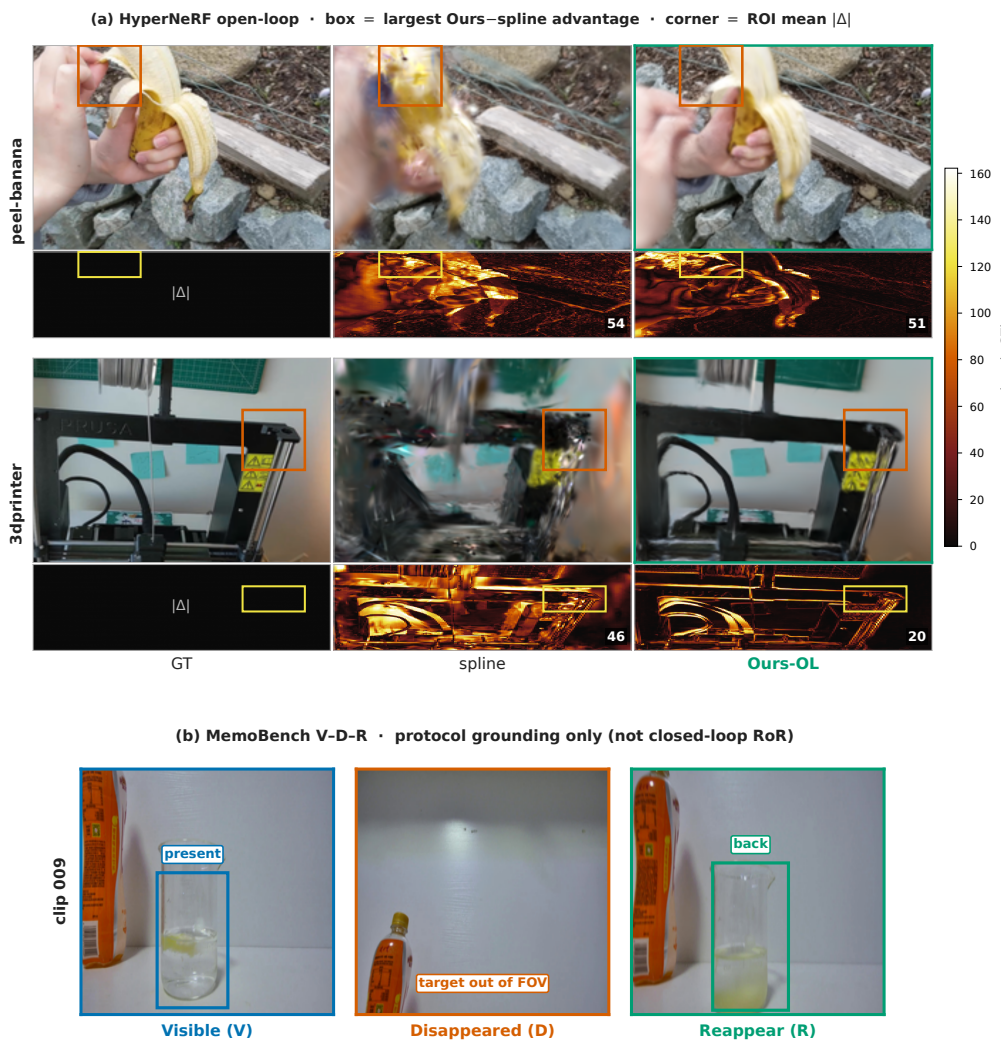

Figure 4: **(a)** HyperNeRF open-loop future frames (GT / analytic spline / Ours-OL) with |pred−GT| strips; vermillion box = largest Ours-vs-spline advantage; corner number = mean $|\Delta|$ inside the box (lower better). **(b)** MemoBench V–D–R protocol grounding: beaker present → out of FOV → back (updated state); *not* closed-loop RoR.

Table 11: Robustness on `fallingball` and `bat` (occ=10, canonical u_p weights, no retraining): relocation RMSE reduction vs. static (%). Top: drift magnitude (× scene extent). Bottom: context fraction.

Drift magnitude (ctx = 0.35)				
drift	0.15	0.30	0.45	0.60
`fallingball`	+44.4	+33.1	+25.4	+17.3
`bat`	+46.6	+43.8	+37.2	+29.7
Reveal density (drift = 0.30)				
ctx_frac	0.25	0.35	0.50	
`fallingball`	+14.6	+33.1	+28.9	
`bat`	+28.1	+43.8	+49.5	

Table 12: NVFi future extrapolation (image PSNR, dB) under a *shared frozen Stage-1*. Same-protocol Stage-2 reimplementations; mixed ranking—backbone sanity check, not a SOTA claim.

Scene	ODE	CfC	GRU	spline	Ours-OL
fallingball	26.15	26.15	18.50	26.37	**30.25**
fan	25.49	25.49	**25.53**	18.01	25.24
bat	**35.60**	35.35	30.25	31.92	34.16
whale	32.04	32.18	29.43	29.37	**36.20**

Table 13: HyperNeRF (Park et al., 2021b) `vrig` future extrapolation (image PSNR, dB), same-protocol reimplementation.

Scene	spline	ODE	Ours-OL
3dprinter	14.52	**16.67**	**16.67**
broom	19.04	**19.28**	19.23
peel-banana	16.93	**18.76**	**18.76**
chicken	**15.45**	11.16	10.60

Table 14: Single arbitrary-time query latency (30k states, ms). CfC is horizon-independent; ODE integrates the path.

Horizon Δt	0.1	0.4	0.8	1.6
CfC (ours)	**1.65**	**1.65**	**2.02**	**1.65**
ODE (torchode)	18.69	17.44	16.91	23.18
Speedup	11.3×	10.6×	8.4×	14.1×